

(12) **UK Patent Application** (19) **GB** (11) **2 384 078** (13) **A**

(43) Date of A Publication 16.07.2003

(21) Application No 0227434.8

(22) Date of Filing 25.11.2002

(30) Priority Data

(31) 2001365801 (32) 30.11.2001 (33) JP

(71) Applicant(s)

Fujitsu Limited
(Incorporated in Japan)
1-1 Kamikodanaka 4-chome, Nakahara-ku,
Kawasaki-shi, Kanagawa 211-8588, Japan

(72) Inventor(s)

Susumu Endo
Yusuke Uehara
Daiki Masumoto
Shuichi Shiitani

(74) Agent and/or Address for Service

Haseltine Lake & Co
Imperial House, 15-19 Kingsway,
LONDON, WC2B 6UD, United Kingdom

(51) INT CL⁷

G06F 17/30

(52) UK CL (Edition V)

G4A AADB

(56) Documents Cited

EP 1262883 A2

US 6243713 B1

US 5598557 A

(58) Field of Search

UK CL (Edition V) **G4A**

INT CL⁷ **G06F**

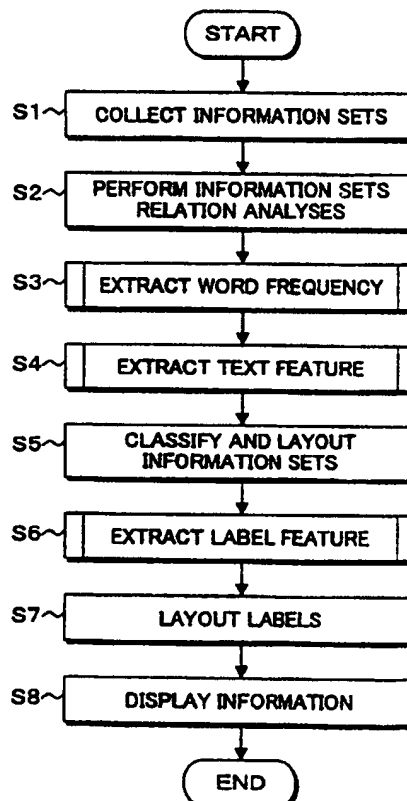
Other: **ONLINE: EPODOC, WPI, PAJ**

(54) Abstract Title

Multimedia information retrieval and display

(57) Paired items of image information and text information correlated to each other are retrieved as information sets. Frequency information on words used in text is extracted from text information in a group of information sets, and text information features are extracted based on frequency information. Text features are used to lay out information sets in a virtual space such that similar pieces of text are located close to each other, and images are displayed at corresponding positions on a screen. Further, important words are extracted from those words extracted from text information in a group of information sets, and those words are laid out in the virtual space in the same manner as with information sets and displayed as labels.

FIG. 3



BEST AVAILABLE COPY

GB 2 384 078

FIG. 1

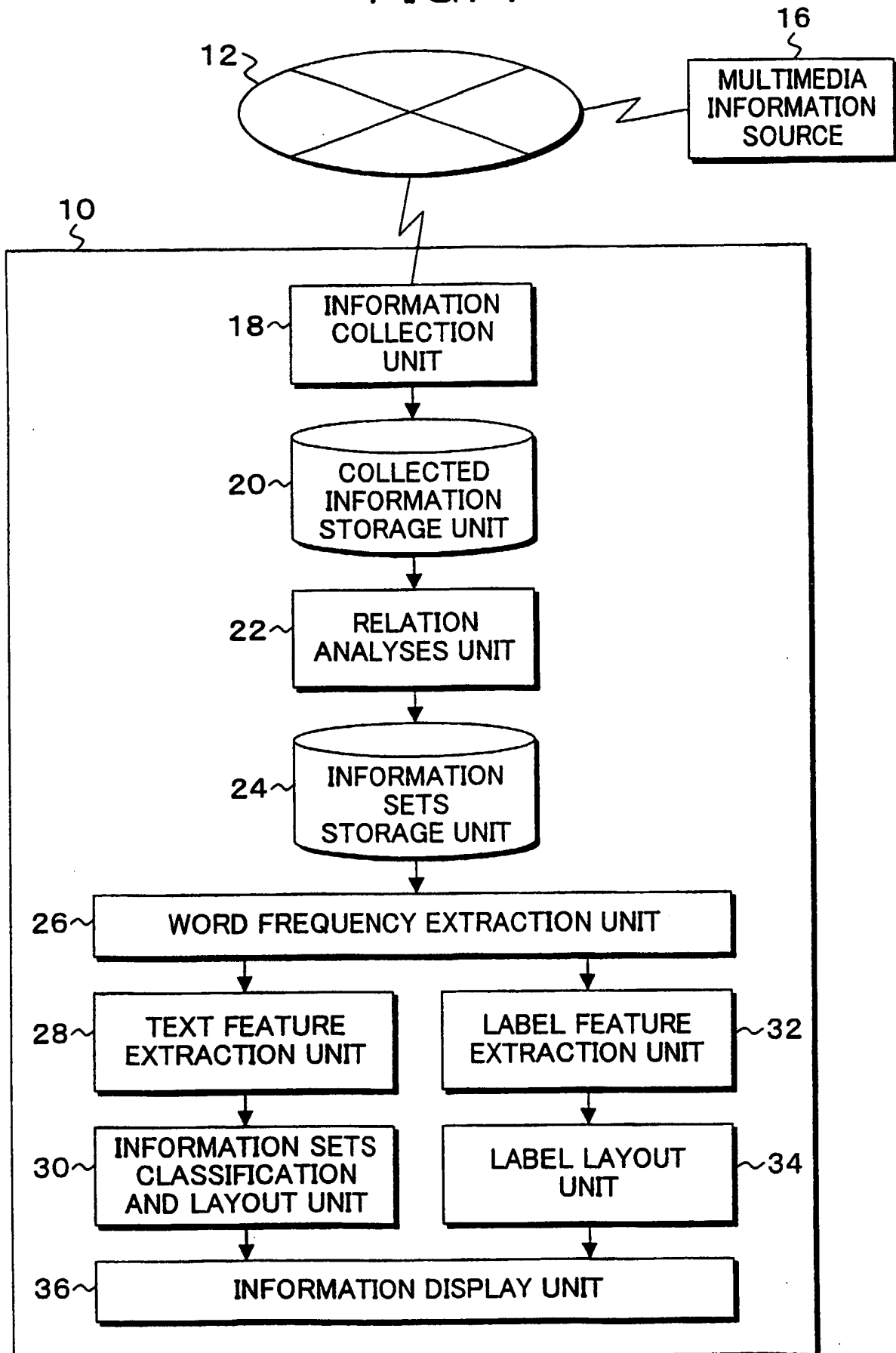


FIG. 2

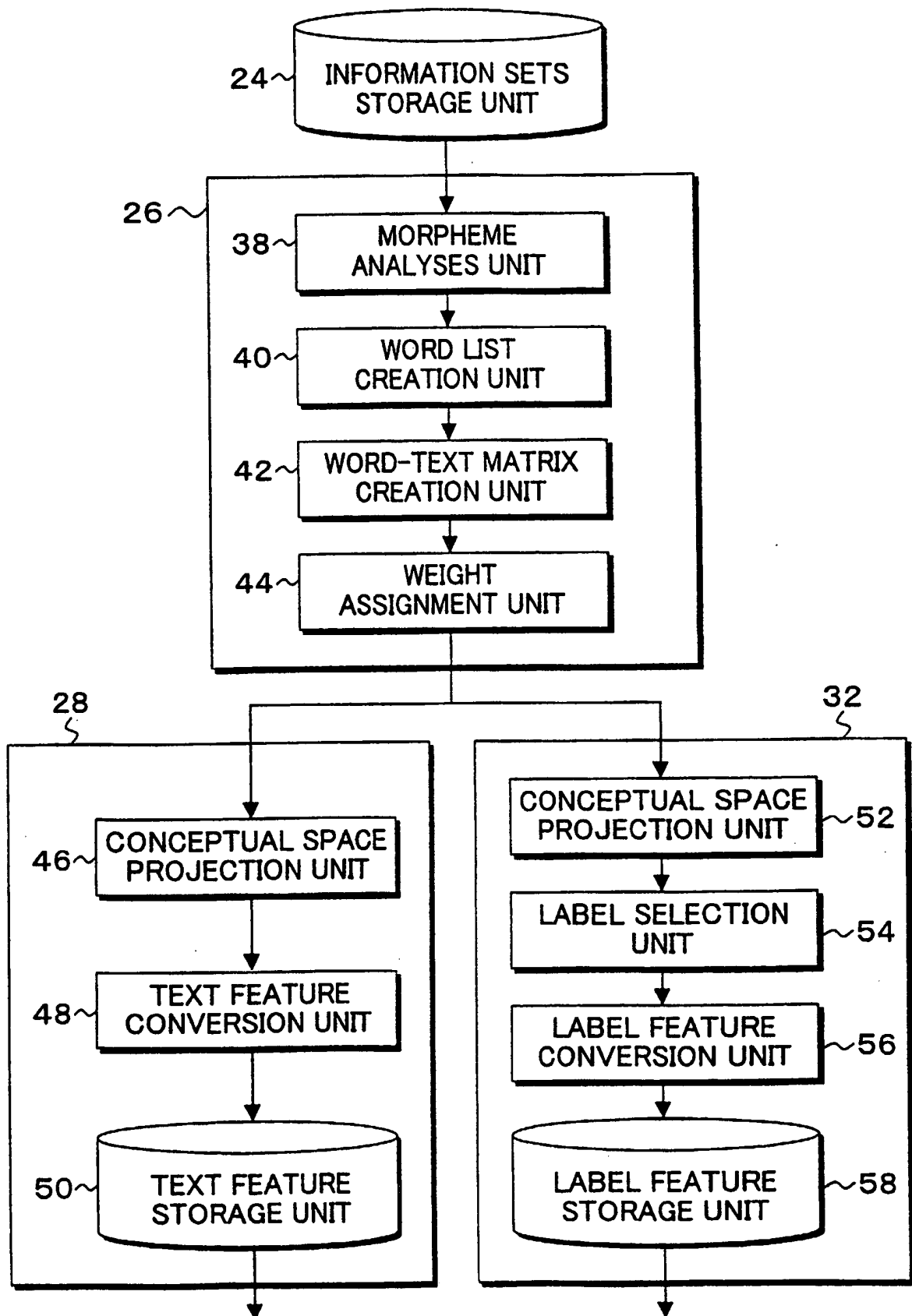


FIG. 3

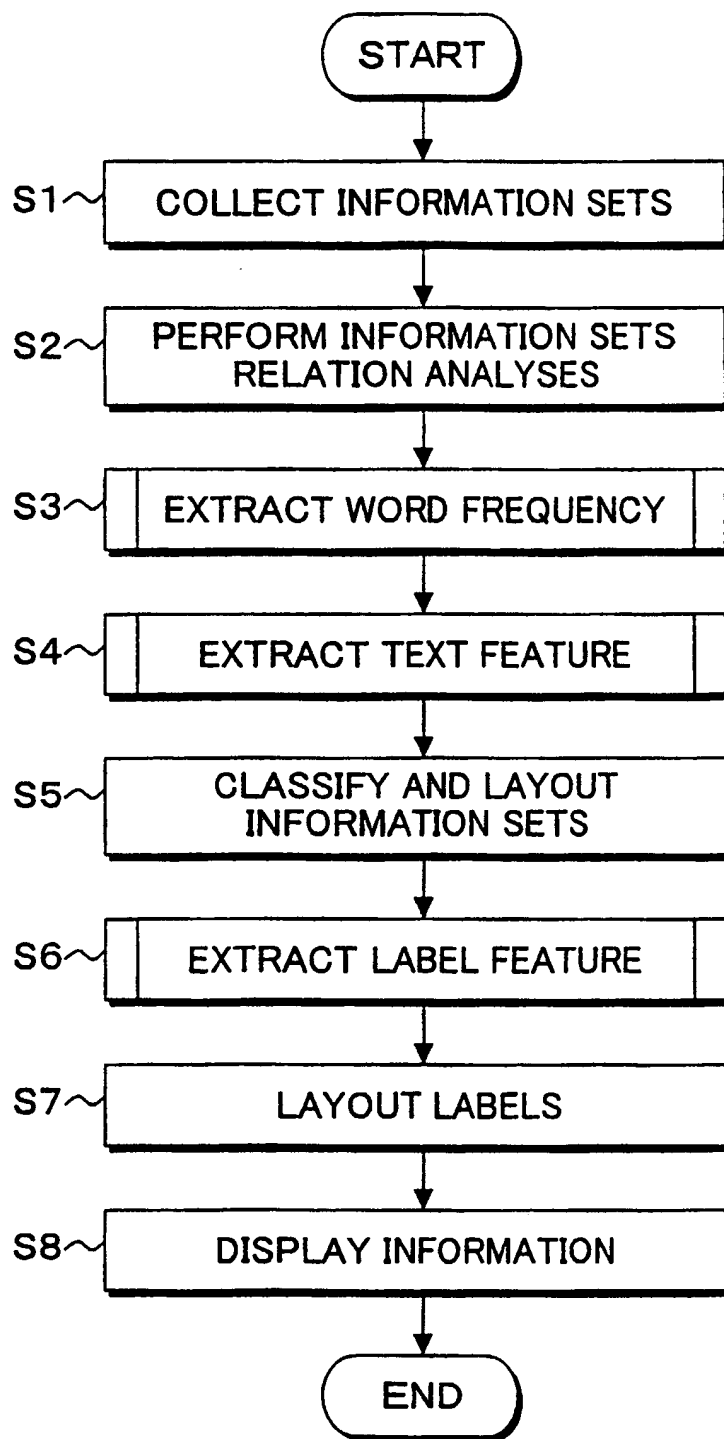


FIG. 4

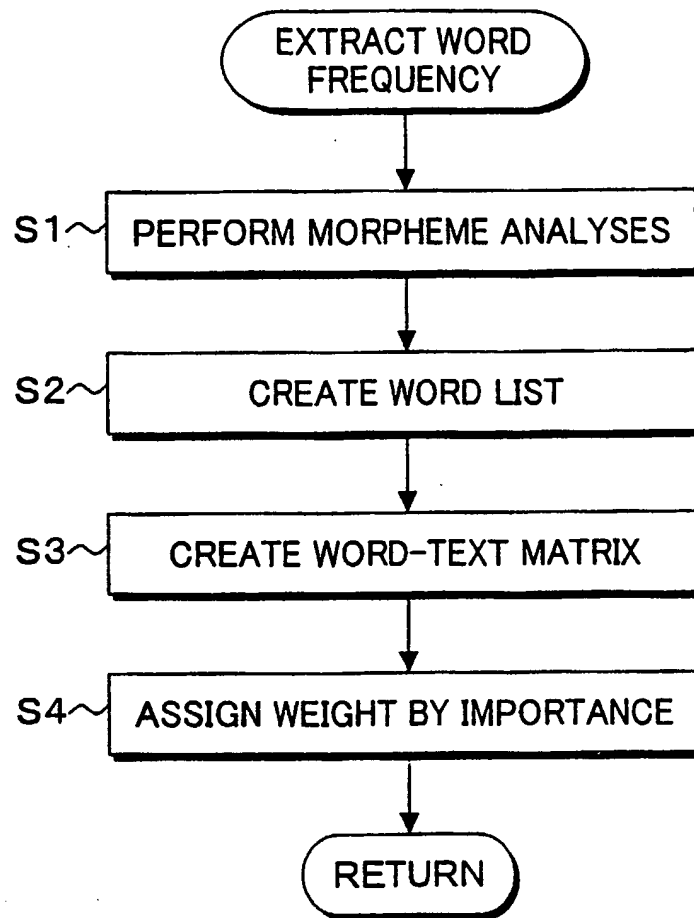


FIG. 5

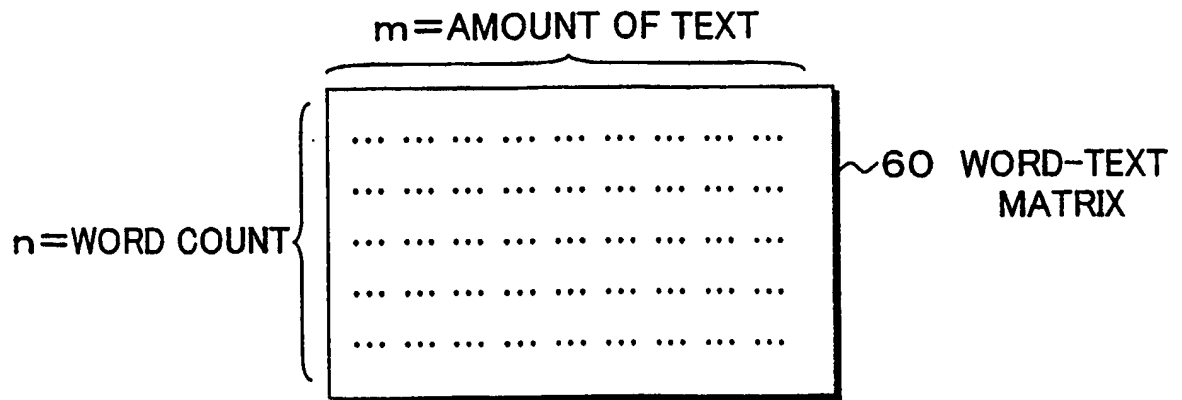


FIG. 6

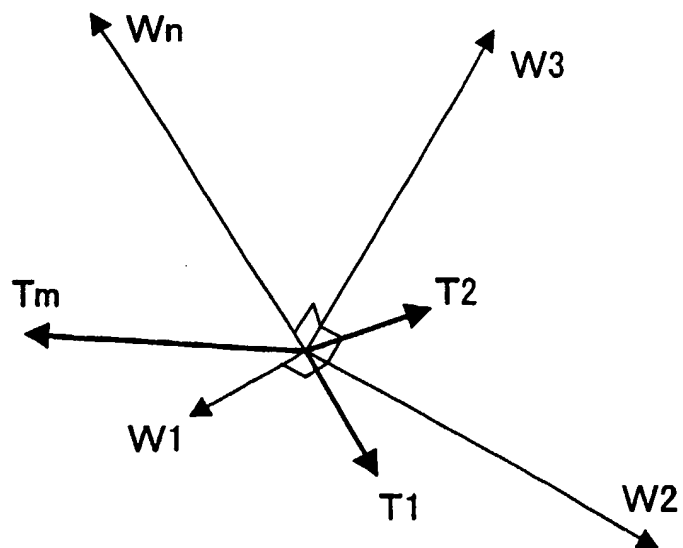


FIG. 7

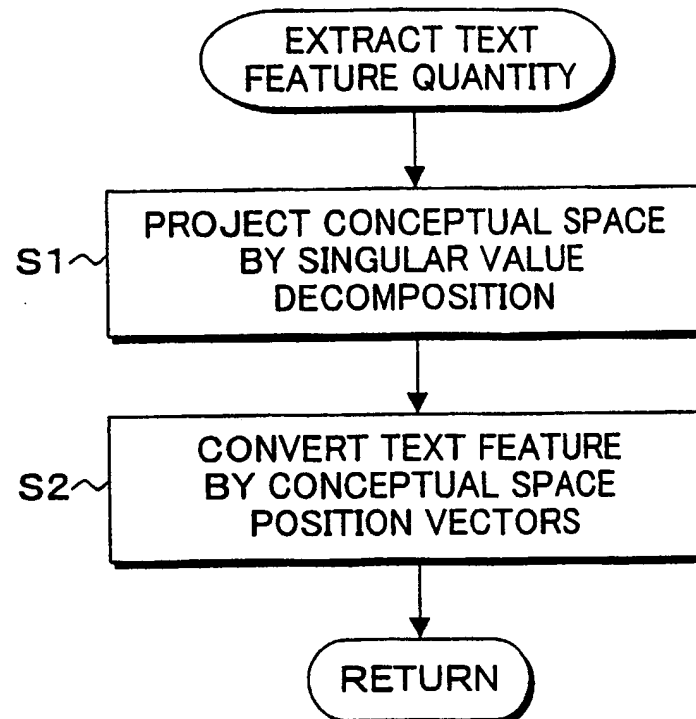


FIG. 8

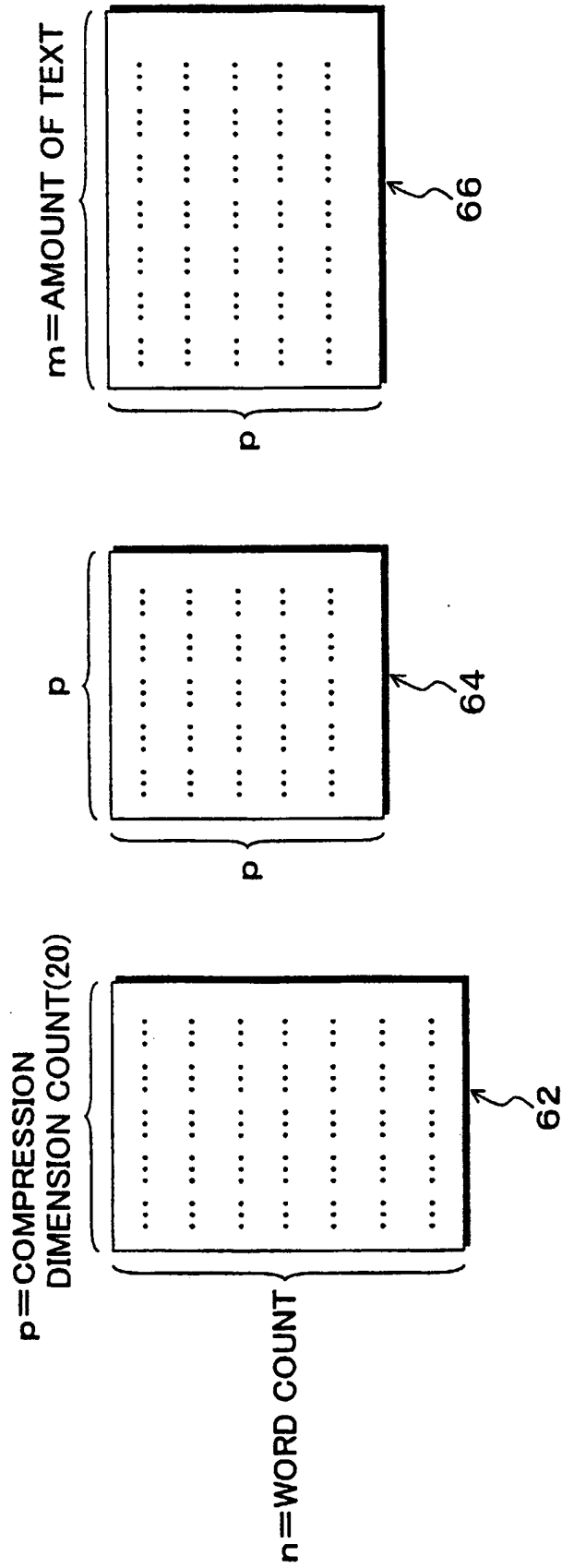
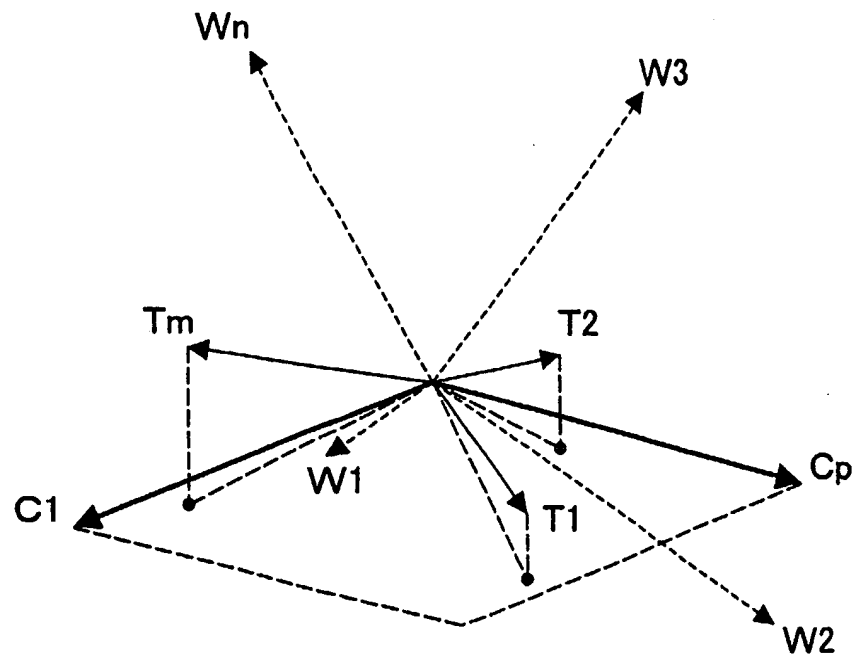


FIG. 9



9/22

FIG. 10

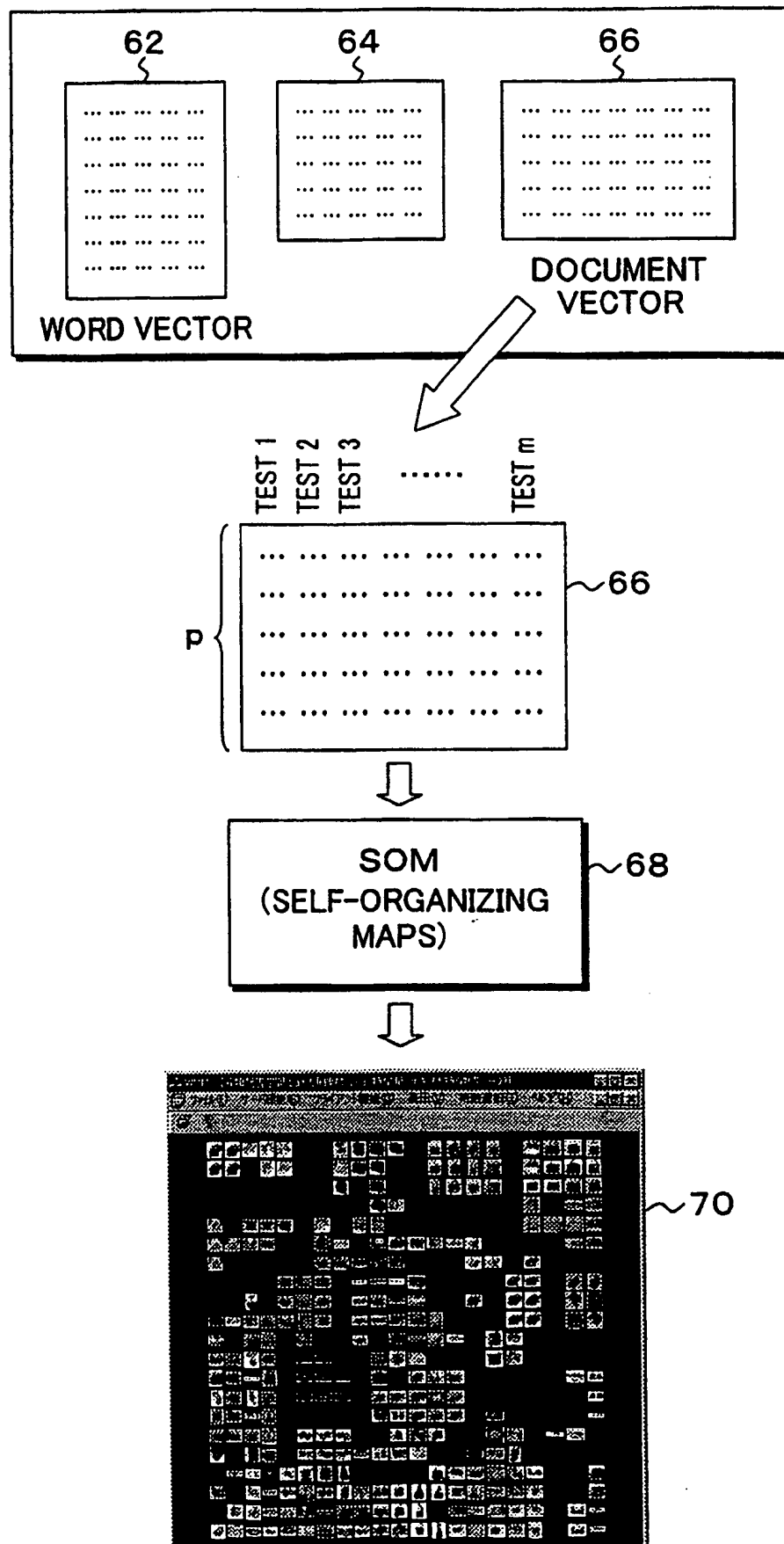


FIG. 11

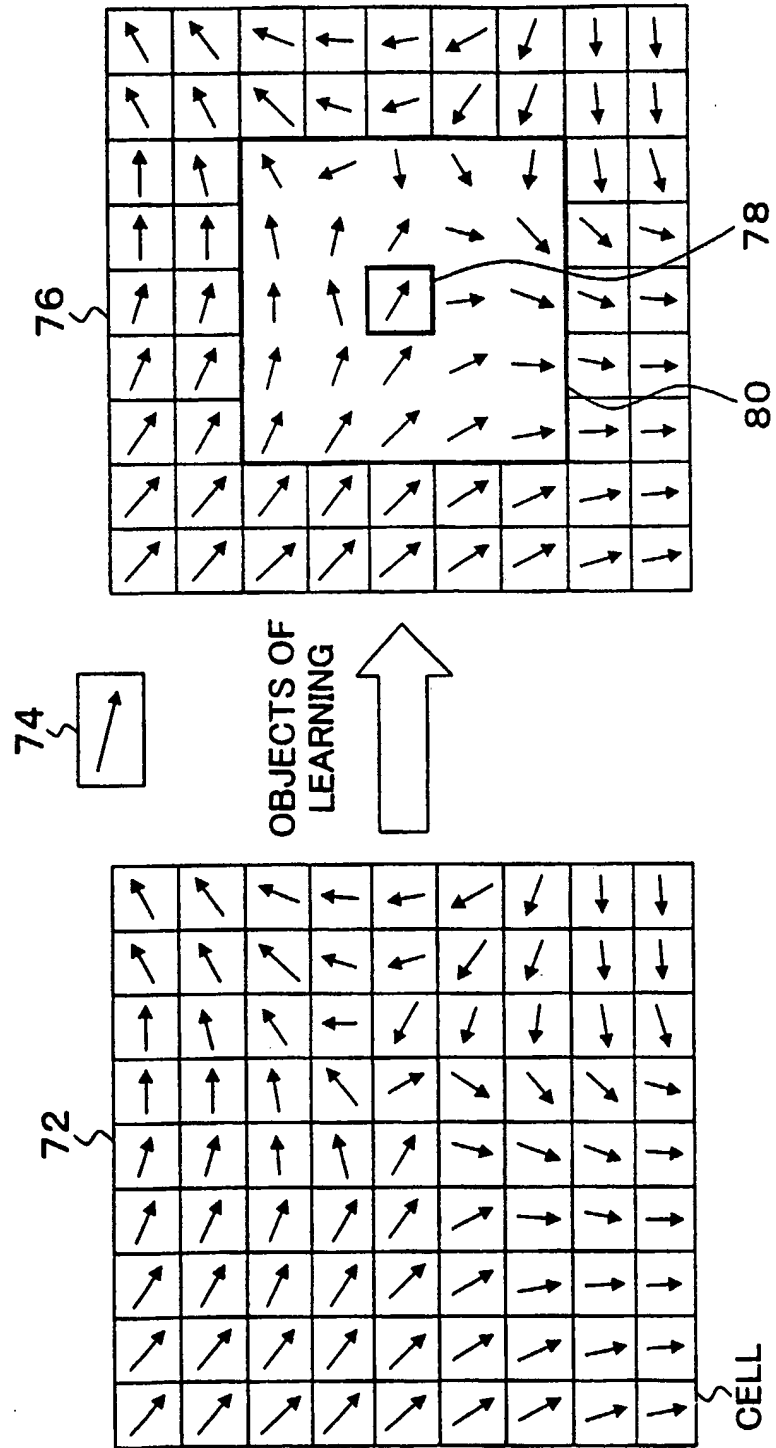


FIG. 12

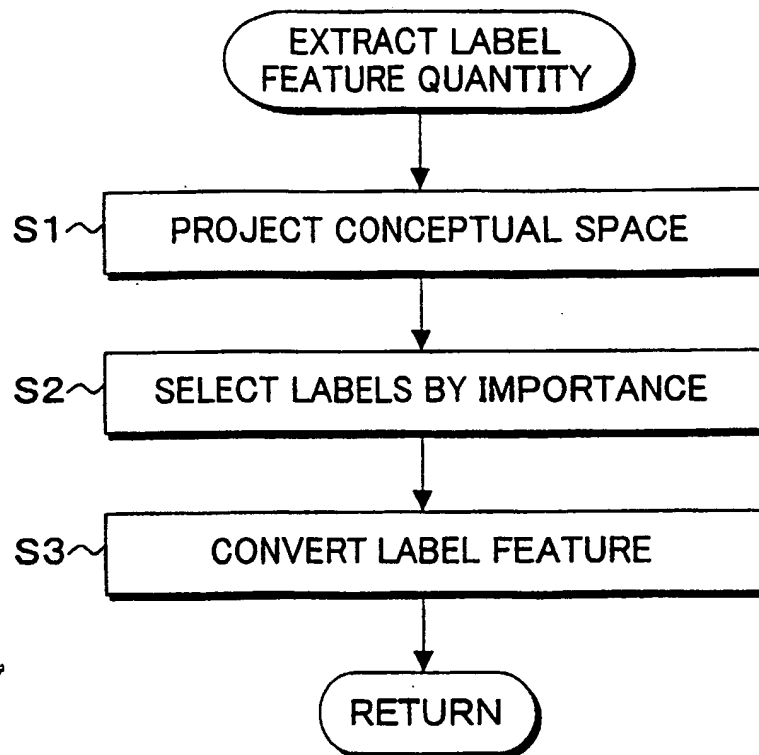
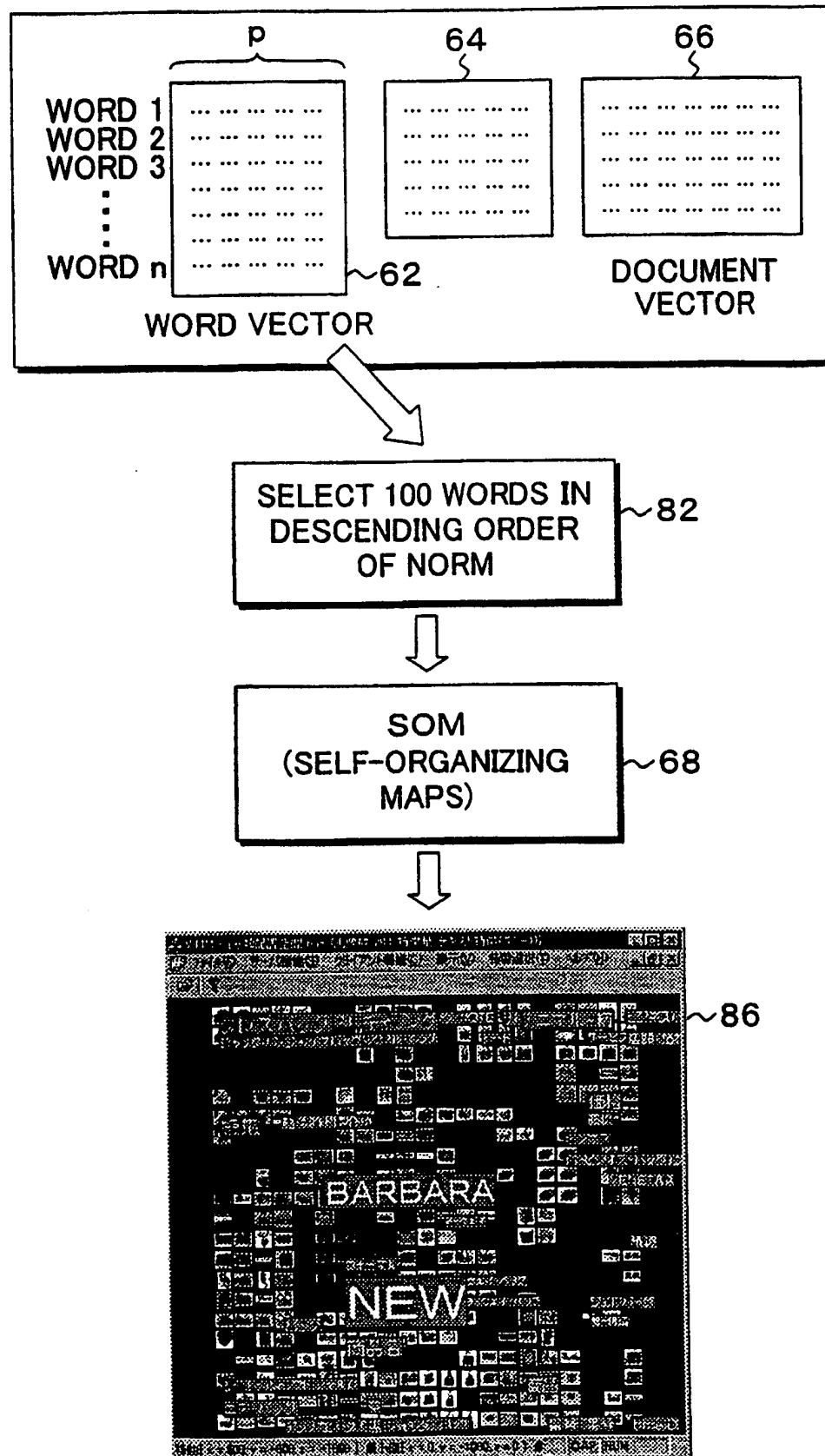


FIG. 13



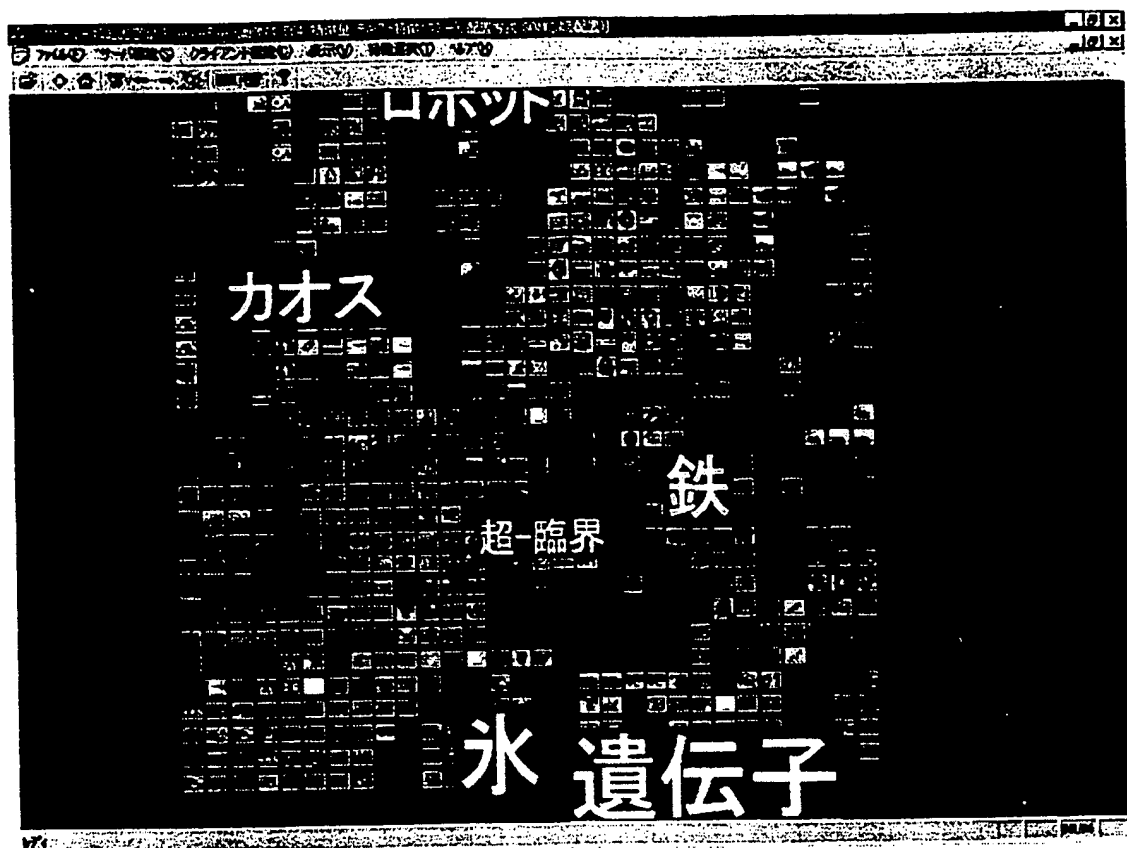


FIG. 14

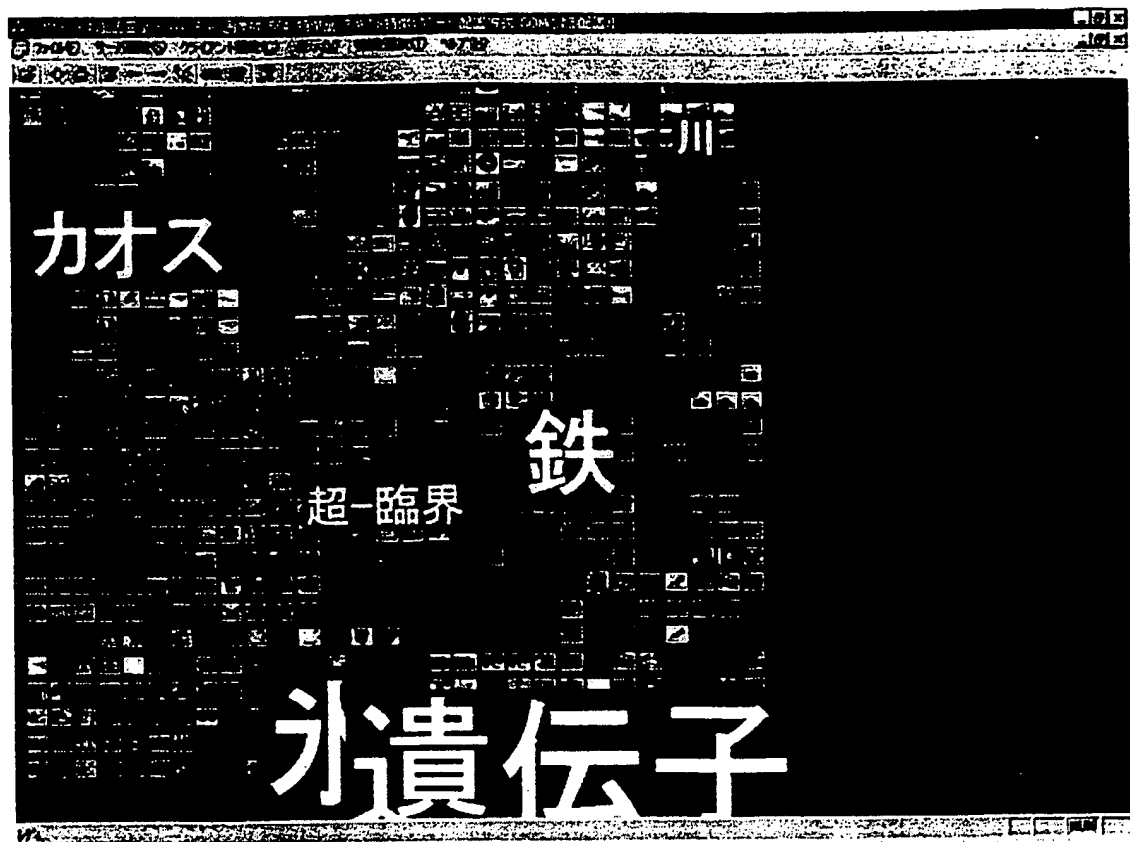


FIG. 15

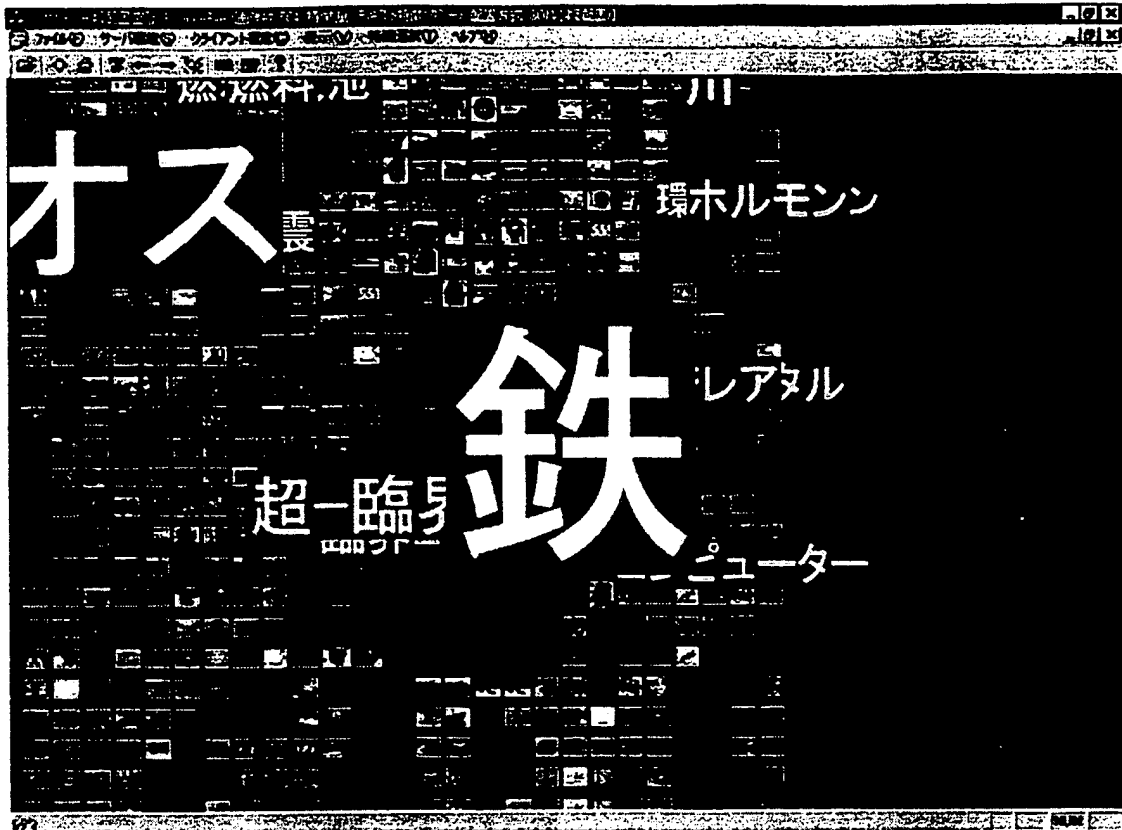


FIG. 16

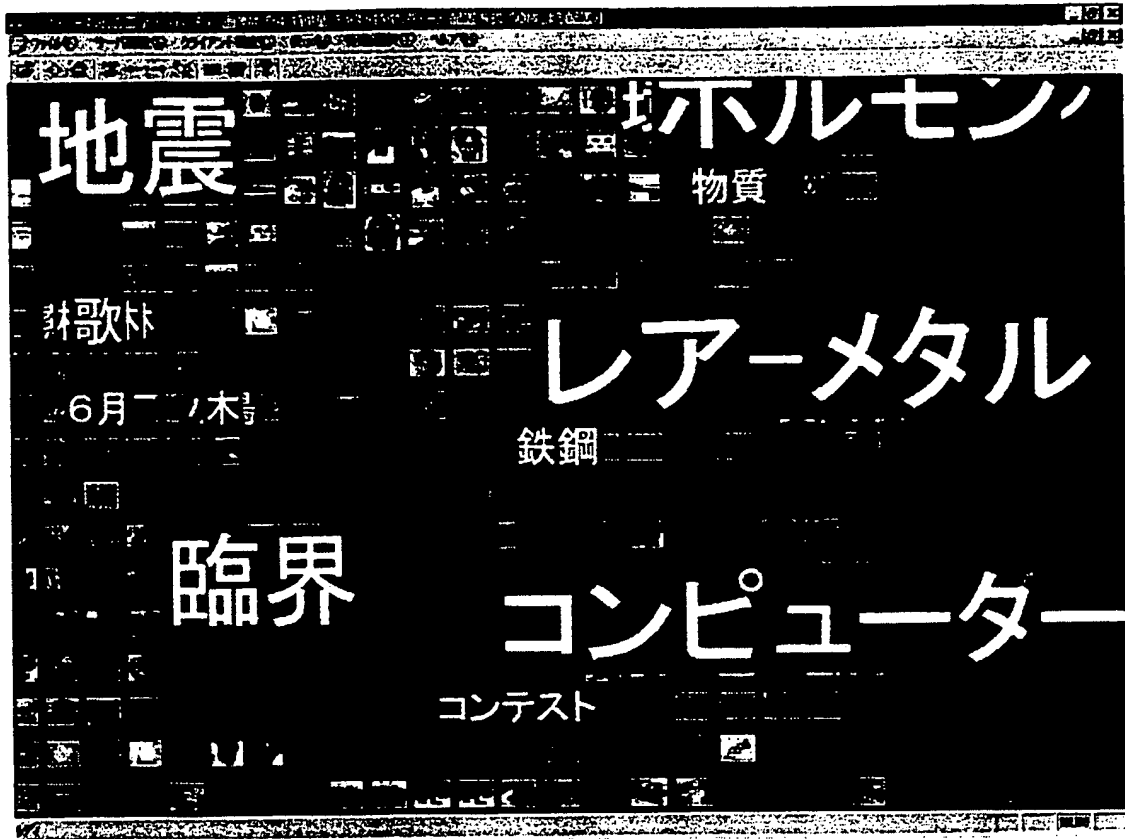


FIG. 17

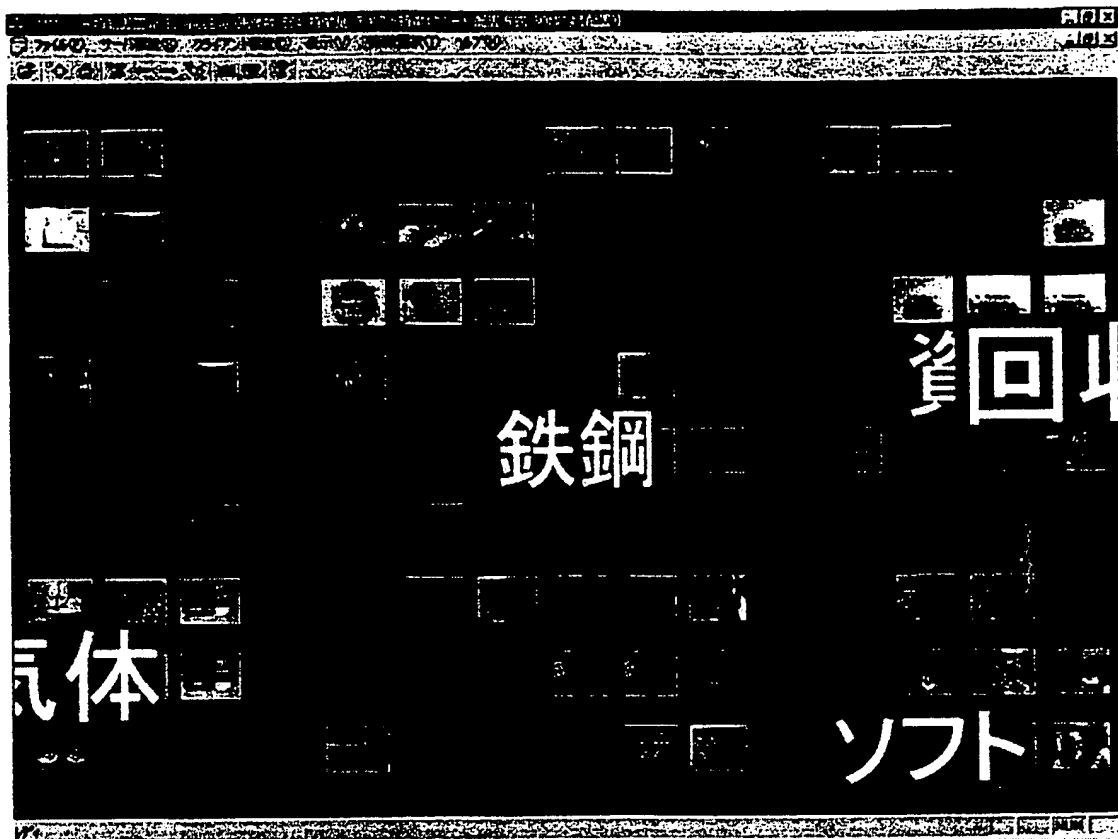


FIG. 18

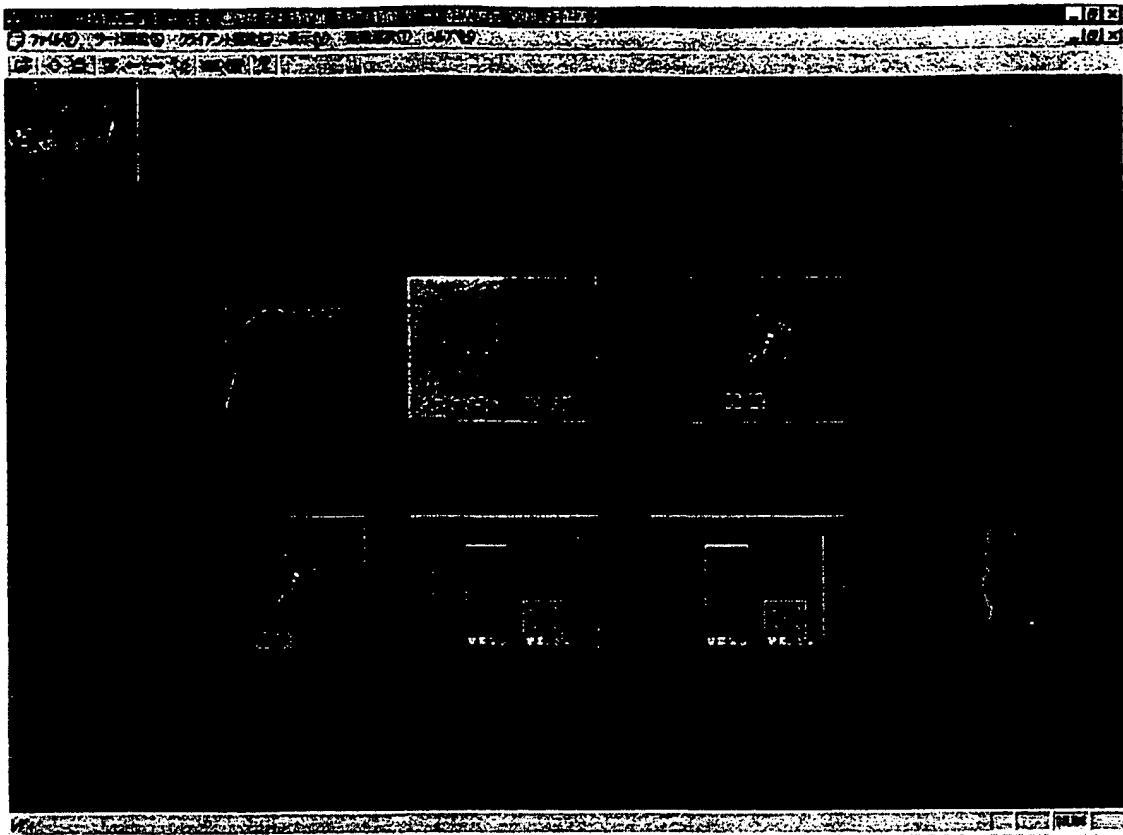


FIG. 19

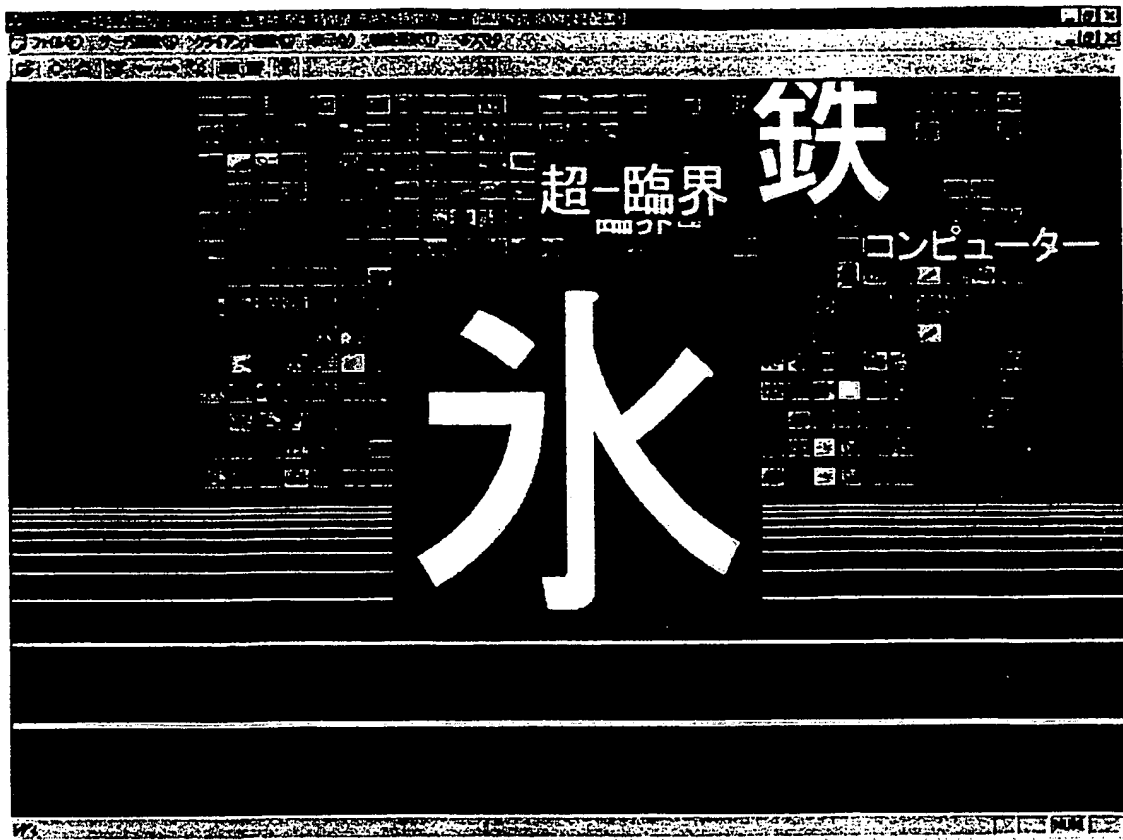


FIG. 20

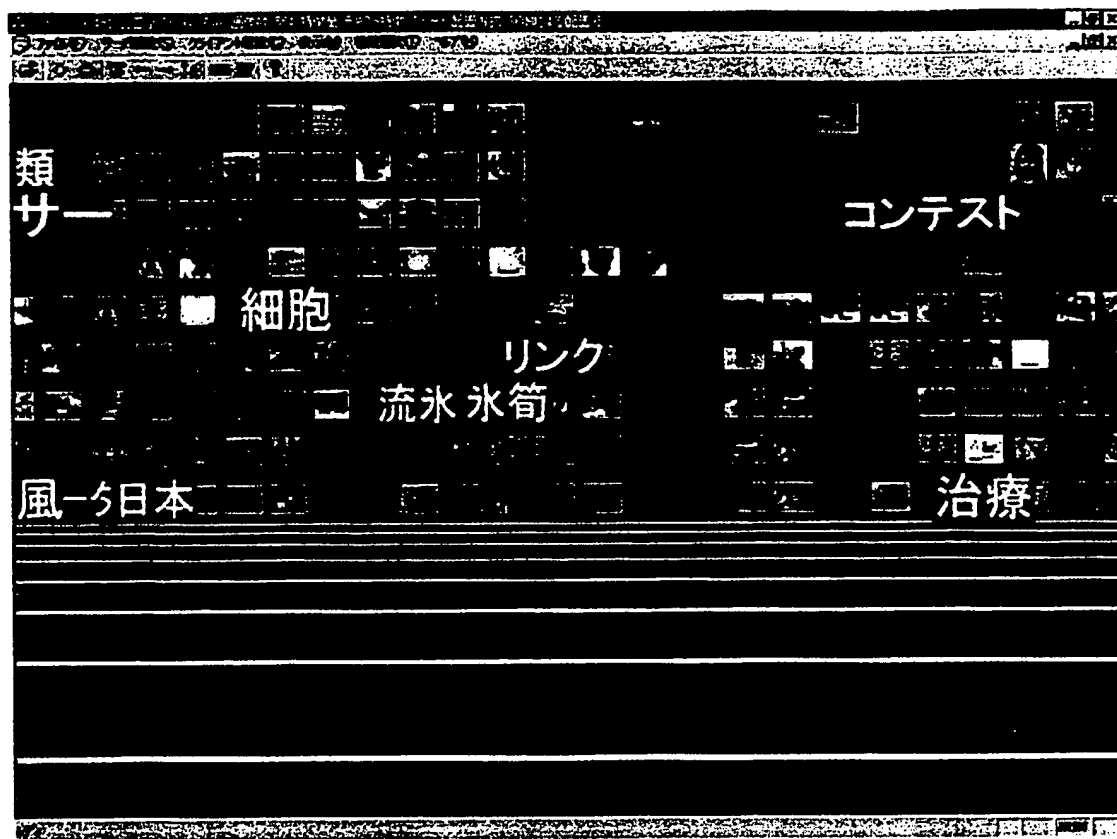


FIG. 21

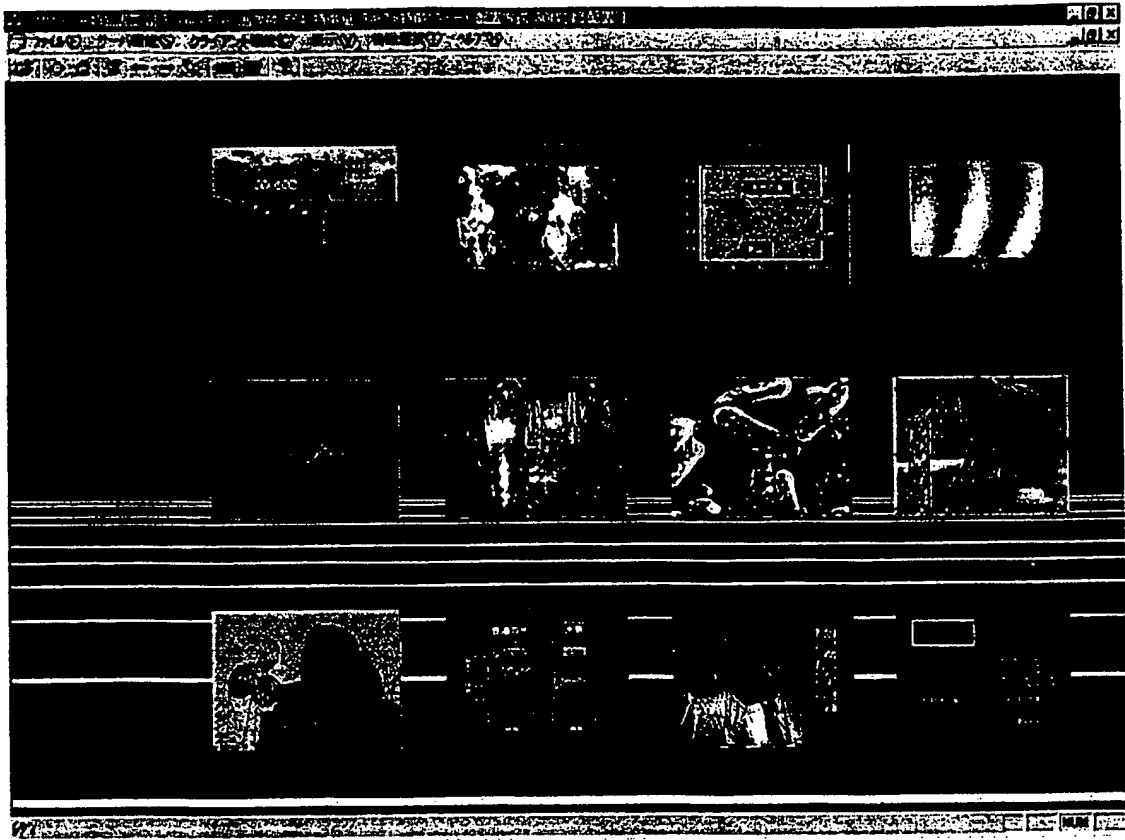
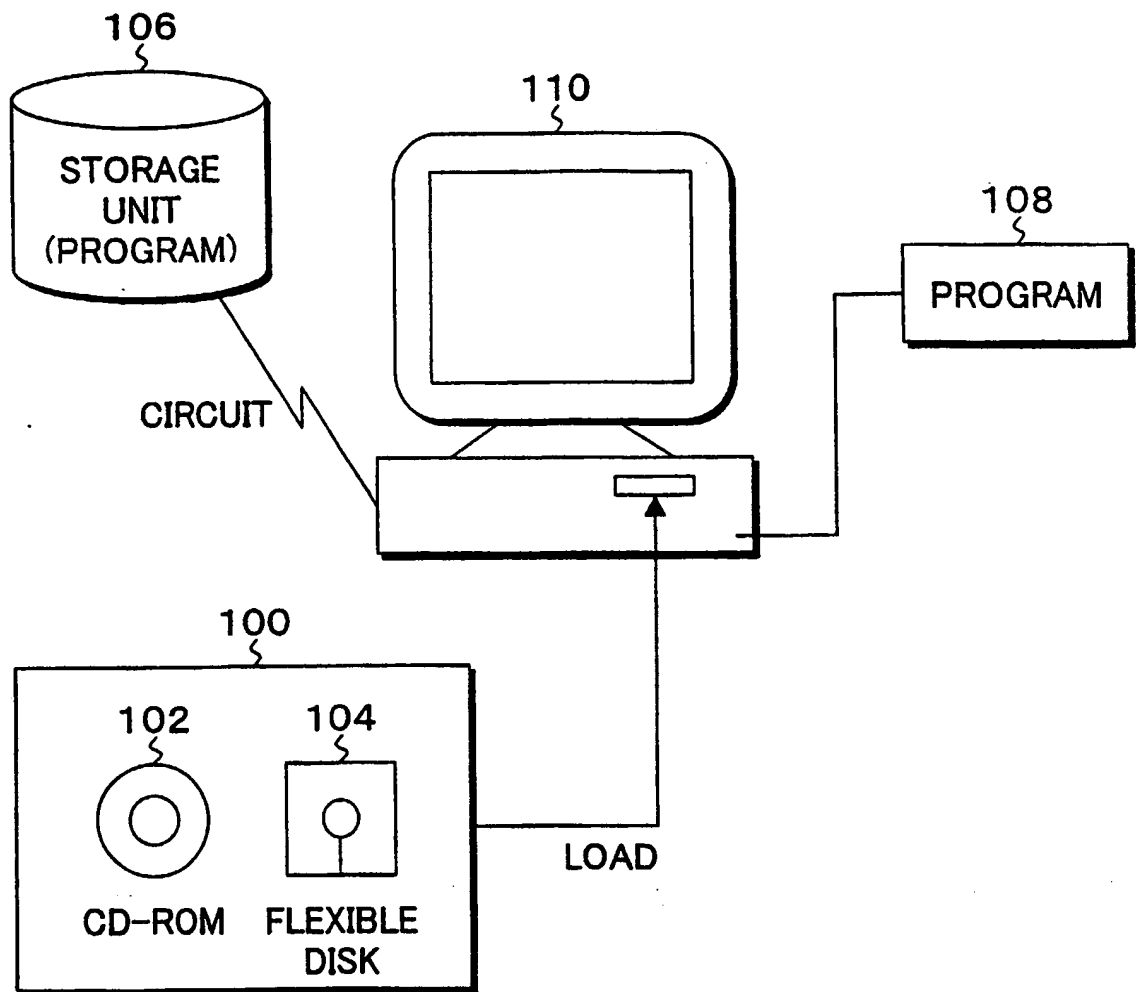


FIG. 22

FIG. 23



MULTIMEDIA INFORMATION RETRIEVAL METHOD,
PROGRAM, RECORD MEDIUM AND SYSTEM

5

The present invention relates generally to a multimedia information retrieval method, program, record medium and system for the efficient retrieval of multimedia information and, more particularly to a multimedia information retrieval method, program, record medium and system for the efficient retrieval of information sets comprised of pairs of image information and text information.

In recent years there have been increasing opportunities for PC users to come into contact with multimedia information due to performance improvements in PCs coupled with growth in the amount of information handled by each user, and as a result it has become necessary to find the required information from among many pieces of multimedia information (documents). There is a conventional method for multimedia information retrieval, in which text

information accompanying multimedia
information is keyword-retrieved.

However, the traditional multimedia
information retrieval method by keyword is
5 unable to assure high search accuracy unless
search criteria precisely representing target
information can be given. In the case of
multimedia information, in particular, it is
difficult to specify search criteria which can
10 precisely represent target information.
Additionally, users themselves have often only
limited understanding of target information.
For this reason, users can only specify broad
search criteria, thus causing a great number
15 of matches to be found. Even if users, faced
with such numerous matches, attempt to narrow
their search by giving a new keyword, they find
it difficult to determine what type of keyword
would be appropriate. Consequently, users
20 need to narrow their search by repeatedly
specifying randomly-chosen keywords, which
has made it difficult to assure high search
accuracy and efficiency.

25

The present invention provides a
multimedia information retrieval method,

program and system which allow highly accurate and efficient retrieval of necessary information from among multimedia information.

5 According to a first aspect of the present invention there is provided a multimedia information retrieval method comprising:

 a word frequency extraction step using as information sets paired image information and
10 text information correlated to each other and extracting constituent word frequency information from the text information within the information sets;

 a text feature extraction step extracting
15 text features, based on the word frequency information of individual information sets;

 an information set classification and layout step classifying and laying out the information sets in a virtual space, based on
20 the text features;

 a label feature extraction step selecting labels from constituent words of the text information within each information set and extracting features of selected labels;

25 a label layout step placing labels at positions corresponding to information sets classified and laid out in the virtual space,

based on the label features; and

an information display step displaying image information and labels of the information sets, placed in the virtual space, depending on the positions of the viewpoint.

For example, the information set classification and layout step includes classifying and laying out the information sets on a two-dimensional plane at a predetermined position in a three-dimensional virtual space, based on the text features, and the label layout step includes placing the labels at the front of the two-dimensional plane in which the information sets are classified and laid out, based on the label features. In this manner, the multimedia information retrieval method of the present invention retrieves as information sets paired image information and text information items correlated with each other. Frequency information on words used in text is extracted from text information in the information sets, and text information features are extracted based on the frequency information. The text features are numerical representations of the text features, so texts similar in content have similar text features. The text features are

used to lay out information sets in a virtual space such that similar pieces of text are located close to each other, and images are displayed at those positions. Thus, while
5 visually grasping the contents by use of images whose contents can rapidly be grasped regardless of large quantities of layout, the user can move to the position where the information exists by walking through the
10 three-dimensional space and examine the image information of the information set in proximity thereof and the text information corresponding thereto to consequently grasp the contents. Further, important words are
15 extracted as labels from those words extracted from text information in the information set, and those labels are laid out in the virtual space in the same manner as with information sets and are displayed as keywords. This label
20 layout enables the user to grasp what information is contained in the information set and where it is localized. For this reason, visual retrieval becomes possible through the display of the images laid out based on the text
25 similarity, whereby efficient retrieval can be carried out regardless of the presence of large quantities of retrieval results. The display

of the labels enables the users to easily grasp what information is contained in the information sets and to easily refer to the relevant information through movement to
5 the vicinity of the label.

The text feature extraction step comprises:

a morpheme analysis step extracting predetermined parts of speech such as nouns,
10 noun compounds (compound nouns) and adjectives by morpheme analyses of text information and creating a word list comprised of words used and their frequencies of occurrence;

a matrix creation step creating a
15 word-text matrix whose rows and columns correspond respectively to text information and words and in which word frequencies of occurrence are laid out as elements; and

a text feature conversion step expressing
20 text information of the word-text matrix by document vectors having coordinate points determined by frequencies of occurrence in a word space which has words as axes, projecting the document vectors onto a low-dimensional
25 space by singular value decomposition and using as text features document vectors representing positions in the low-dimensional

space.

The matrix creation step of the text feature conversion step comprises a weight assignment step assigning weight to elements
5 in the word-text matrix, based on word frequencies of occurrence in each text. This allows words appearing in specific texts only to be handled as being higher in degree featuring that text, i.e., higher in
10 significance than words appearing in all texts without exception.

The label feature extraction step comprises a label selection step figuring out the significance of each constituent word of
15 text information within each of the information sets and selecting words to be used as labels, based on the significance figured out.

The label selection step comprises:
20 a morpheme analysis step extracting nouns and noun compounds by morpheme analyses of text information and creating a word list comprised of words used and their frequencies of occurrence;

25 a matrix creation step creating a word-text matrix whose rows and columns correspond respectively to text information

and words and in which the word frequencies of occurrence are laid out as elements; and a label feature conversion step expressing words of the word-text matrix by word vectors
5 having coordinate points determined by frequencies of occurrence in a text space which has individual pieces of text information as axes, projecting the word vectors onto a low-dimensional space by singular value
10 decomposition and using as label features word vectors representing positions in the low-dimensional space; wherein a predetermined number of words are selected as labels in descending order of significance
15 which is represented by lengths of the word vectors. The word-text matrix in this label selection step is the same as one created in the text feature extraction step, and hence use of the same word-text matrix allows omission
20 of the morpheme analysis step and the matrix creation step. The label layout step includes displaying labels such that the higher the significance determined by the label selection step, the more the labels are
25 displayed toward the front of the virtual space. The information display step includes changing how labels appear and the size in which they

appear, depending on the position of the viewpoint relative to the virtual space. The information display step also includes fixing the horizontal position of the label relative to
5 image information regardless of a horizontal displacement of the viewpoint in the virtual space and changing how labels appear and the size that they appear depending on a change in the position of the viewpoint in the
10 direction of depth. Thus, in walking through the virtual space and with movement of the viewpoint, the position of the label relative to the image is unvaried regardless of horizontal movement of the viewpoint,
15 preventing labels closer to the forefront from moving horizontally to a larger extent, making the correlation with the image unclear, as happens in the case of execution of the ordinary three-dimensional coordinates calculation.

20 The multimedia information retrieval method of the first aspect of the present invention further comprises an information collection step collecting, for example, from the Internet, information sets comprised of paired image and text information correlated to each other. In this case, the
25 information collection step comprises a relation analysis step analyzing the relationship

between image information and text information and determining the range of information to be collected as information sets, if the relationship between the image information and the text information is unclear.

According to a second aspect of the present invention there is provided a program for retrieving multimedia information. This program allows a computer to execute:

10 a word frequency extraction step, using as information sets paired image information and text information correlated to each other, and extracting constituent word frequency information from the text information within
15 the information sets;

a text feature extraction step extracting text features, based on the word frequency information of individual information sets;

an information set classification and
20 layout step classifying and laying out the information sets in a virtual space, based on the text features;

a label feature extraction step selecting labels from constituent words of the text
25 information within each information set and extracting features of selected labels;

a label layout step placing labels at

positions corresponding to information sets classified and laid out in the virtual space, based on the label features; and

an information display step displaying
5 image information and labels of the information sets, placed in the virtual space, depending on the position of the viewpoint.

According to a third aspect of the present invention there is provided a computer
10 readable record medium having therein stored a program for retrieving multimedia information. The program stored in this record medium allows a computer to execute:

a word frequency extraction step using as
15 information sets paired image information and text information correlated to each other wherein constituent word frequency information is extracted from the text information within the information sets;

20 a text feature extraction step extracting text features, based on the word frequency information of individual information sets;

an information set classification and layout step classifying and laying out the
25 information sets in a virtual space, based on the text features;

a label feature extraction step selecting

labels from constituent words of the text
information within each information set and
extracting features of said selected labels;

5 a label layout step placing labels at
positions corresponding to information sets
classified and laid out in the virtual space,
based on the label features; and

an information display step displaying
image information and labels of the
10 information sets, placed in the virtual space,
depending on the position of the viewpoint.

According to a fourth aspect of the
present invention there is provided a
multimedia information retrieval system.

15 This system comprises:

a word frequency extraction unit using, as
information sets, paired image information and
text information correlated to each other, and
extracting constituent word frequency
20 information from the text information within
the information sets;

a text feature extraction unit extracting
text features, based on the word frequency
information of individual information sets;

25 an information set classification and
layout unit classifying and laying out the
information sets in a virtual space, based on

the text features;

a label feature extraction unit selecting labels from constituent words of the text information within each information set and
5 extracting features of the selected labels;

a label layout unit placing labels at positions corresponding to information sets classified and laid out in the virtual space, based on the label features; and

10 an information display unit displaying image information and labels of the information sets, placed in the virtual space, depending on the position of the viewpoint.

The present invention also embraces a program for
15 implementing the above method, which may be stored on a record medium.

The above and other aspects, features and advantages of the present
20 invention will become more apparent from the following detailed description made with reference, by way of example only, to the accompanying drawings in which:

Fig. 1 is a block diagram representing the
25 functional configuration of an embodiment of the invention;

Fig. 2 is a block diagram representing

detailed configurations of a word frequency extraction unit, a text feature extraction unit and a label feature extraction unit;

Fig. 3 is an example flowchart depicting
5 multimedia information retrieval according to an embodiment;

Fig. 4 is a flowchart of the word frequency extraction shown in Fig. 3;

Fig. 5 is an explanatory drawing of a
10 word-text matrix created by word frequency extraction shown in Fig. 3;

Fig. 6 is an explanatory drawing representing a word-text matrix by document vectors in an n -dimensional word space;

15 Fig. 7 is a flowchart of classification and layout of information sets shown in Fig. 3;

Fig. 8 is an explanatory drawing showing singular value decomposition of a word-text
20 matrix;

Fig. 9 is an explanatory drawing of a low-dimensional space converted from Fig. 6 by singular value decomposition;

Fig. 10 is an explanatory drawing of
25 classification and layout of information sets shown in Fig. 3;

Fig. 11 is an explanatory drawing of

layout by self-organizing maps (SOM);

Fig. 12 is a flowchart of label feature extraction shown in Fig. 3;

Fig. 13 is an explanatory drawing of label
5 layout of this invention;

Fig. 14 is an explanatory drawing of a search screen according to this invention;

Fig. 15 is an explanatory drawing of the search screen seen when the viewpoint is moved
10 toward label "鉄" in the direction of depth;

Fig. 16 is an explanatory drawing of the search screen seen when the viewpoint is moved further toward label "鉄" in the direction of depth;

15 Fig. 17 is an explanatory drawing of the search screen seen when the viewpoint moves past label "鉄";

Fig. 18 is an explanatory drawing of the search screen seen when the viewpoint is moved
20 further toward label "鉄鋼" in the direction of depth;

Fig. 19 is an explanatory drawing of the search screen seen when the viewpoint moves past label "鉄鋼";

25 Fig. 20 is an explanatory drawing of the search screen seen when the viewpoint is moved toward label "氷" in the direction of depth;

Fig. 21 is an explanatory drawing of the search screen seen when the viewpoint moves past label "冰";

Fig. 22 is an explanatory drawing of the search screen seen when the viewpoint moves past label "冰箭" shown in Fig. 21; and

Fig. 23 is an explanatory drawing of the embodiment of a record medium in which the program is stored.

10 An embodiment of the present invention will now be discussed regarding the accessing of multimedia information on the Internet. However the various embodiments of the invention are not solely limited to accessing information on line and can be used to access multimedia information on
15 networks, servers or databases as appropriate.

Fig. 1 is a block diagram depicting the system configuration of a multimedia information retrieval system embodying this invention, and the multimedia information retrieval system 10 according to this embodiment accesses a
20 multimedia information source 16, comprised for example of a plurality of web pages, via the Internet 12 and treats information sets comprised of pairs of correlated items of image and text information within collected multimedia information as target information. That is, in this example, the multimedia
25 information retrieval system 10 collects information sets comprised of pairs of image and text information from the multimedia information source 16 via the Internet 12.

It then extracts text features from the text information contained in the collected information sets and lays out information sets in a virtual space based on similarities between different pieces of text information, before displaying image information. Further, the system displays labels each of which will serve as a landmark for retrieving image information to indicate the contents and locations of individual pieces of image information. In order to ensure that such images and labels of multimedia information can be displayed and retrieved three-dimensionally, the multimedia information retrieval system 10 comprises an information collection unit 18, a collected information storage unit 20, a relation analysis unit 22, an information sets storage unit 24, a word frequency extraction unit 26, a text feature extraction unit 28, an information set classification and layout unit 30, a label feature extraction unit 32, a label layout unit 34 and an information display unit 36. Of these units, the word frequency extraction unit 26 handling information set classification and layout, the text feature extraction unit 28 and the label feature

extraction unit 32 are depicted in detail in a block diagram shown in Fig. 2.

The word frequency extraction unit 26 shown in Fig. 2 comprises a morpheme analysis unit 38, a word list creation unit 40, a word-text matrix creation unit 42 and a weight assignment unit 44. The text feature extraction unit 28, on the other hand, comprises a low-dimensional space projection unit 46, a text feature conversion unit 48 and a text feature storage unit 50. Moreover, the label feature extraction unit 32 comprises a low-dimensional space projection unit 52, a label selection unit 54, a label feature conversion unit 56 and a label feature storage unit 58.

Fig. 3 is a flowchart of the program run by (or the method carried out by) the system shown in Fig. 1, and Steps 1 through to 8 correspond to the functions of the respective processing units of the system. In Step S1, information sets are collected by the information collection unit 18. In Step S2, information sets relation analyses are performed by the relation analyses unit 22. In Step S3, word frequency is extracted by the word frequency extraction unit 26. In Step S4, text

feature is extracted by the text feature extraction unit 28. In Step S5, information sets are classified and laid out by the information set classification and layout unit.

5 Steps S1 to S5 are designed to classify and lay out information sets comprised of pairs of image and text information in this example collected over the Internet, based on similarities between information sets. In
10 the following Step 6, label feature extraction and label selection are performed by the label feature extraction unit 32, based on the frequency (significance) of words extracted by the word frequency extraction
15 unit 26. In Step S7, labels are laid out by the label layout unit 34. The processing performed in these Steps S6 and S7 is designed to classify and lay out labels for use as keywords for information sets laid out in
20 a virtual space. In the final Step S8, image information within information sets, classified and laid out in a three dimensional display space, and corresponding labels are displayed by the information display unit 36.
25 Next, the processing performed by the multimedia information retrieval system 10 shown in

Figs. 1 and 2, is described in detail in accordance with the flowchart shown in Fig. 3. To collect information sets in Step S1 shown in Fig. 3, the Kloro (a crawler) known as an Internet information collection robot is employed. The Kloro, which functions as the information collection unit 18 in Fig. 1, searches over the Internet 12 in accordance with specified criteria and also tracks links, thereby tracking a plurality of web pages as the multimedia information source 16. This allows the Kloro to collect images and their related text on WEB pages and to store image-text pairs as information sets. There are two methods to specify criteria with the Kloro. One of them is to use URLs, by which pages under and linked to the specified URL are searched. Another method is to pass the keyword to the text search server (text search engine) first and then use a URL list returned in response to the keyword for searching WEB pages. Thus, information including image-text pairs, collected by the information collection unit 18 which performs collection of information sets in Step S1, is stored in the collected information storage unit 20. In the following Step S2, the relation analysis unit 22 performs information

sets relationship analyses, generates information sets comprised of image-text pairs and stores these in the information sets storage unit 24. Information sets relationship analyses by the
5 relation analysis unit 22 are made by choosing image-text pairs whose relationship is unclear from among those collected from the Internet, analyzing their relationship with the user, determining the range of information sets and
10 storing those sets which fall within the range.

These relationship analyses are performed by analyzing Kloria-collected HTML files, estimating text-image relationship
15 based, for example, on the number of line feeds inserted between image and text and also on the HTML tags and excerpting a predetermined range of text in order to determine information sets. In the succeeding Step S3, word frequency
20 extraction is performed by the word frequency extraction unit 26.

This word frequency extraction comprises morpheme analyses in Step S1, word list creation in Step S2, word-text matrix creation
25 in Step S3 and weight assignment by significance in Step S4 as shown in Fig. 4 as the subroutine's flowchart of Step S3. This

processing corresponds to the functional configuration of the word frequency extraction unit 26 shown in Fig. 2. Word frequency extraction in Fig. 4 starts with morpheme analysis in Step S1. In morpheme analysis, important words and their frequencies are extracted from text. More specifically, the morpheme analysis method is used to analyze linguistic information in text and separate it into morphemes. Further, of the separated morphemes those for parts of speech such as articles, conjunctions and pronouns which are unnecessary or do not occur frequently are deleted. In an embodiment of the invention, only nouns are extracted by these morpheme analyses. Moreover, specific morphemes are deleted or substituted based on rules. Further, whether a plurality of nouns can be combined, based on frequencies and rules, to form a compound noun representing a single meaning is evaluated, and if possible, a compound noun is generated by combining a plurality of nouns. Such morphemes, namely, nouns or compound nouns obtained by morpheme analysis are hereafter referred to as words. By performing such morpheme analysis over the entire text, a word list is created in which

all the words, used in the entire text in Step S2, are arranged. In the following Step S3, a word-text matrix 60, in which rows and columns correspond respectively to individual pieces of text and words, is created as shown in Fig. 5 based on the entire text's word list. Each of the elements in this word-text matrix 60 represents the frequency of each word in the text. Consequently, each element in the word-text matrix 60 can be expressed in vector form as shown in Fig. 6. The coordinate space in Fig. 6 represents document vectors T_1, T_2, \dots, T_m , whose lengths are equal to word frequencies and which correspond to text 1, text 2, ..., text m arranged along rows, in a word coordinate space having coordinate axes W_1, W_2, \dots, W_n which correspond to word 1, word 2, ..., word n arranged along columns in the word-text matrix 60 shown in Fig. 5. In the succeeding Step S4 shown in Fig. 4, weight is assigned by significance to frequencies, each of which is an element in the word-text matrix in Fig. 5. In considering the significance of each word in each piece of text, it is generally possible to regard words which occur in only part of the text as being more important than those appearing uniformly in the entire text since the former are more useful

for identifying that text. For this reason, in assigning weight in Step S4, TFIDF (Inverse Document Frequency of Term Frequency) is used to assign weight to the frequency of each word which is an element of the word-text matrix 60. This weight assignment by TFIDF is performed such that the lower the probability of occurrence of a word in other text, the higher the significance of that word. More specifically, weight $W(t_i)$ assigned to the frequency of each word is given by the following formula:

$$IDF(t_i) = \log\{(\text{total text amount } m) / (\text{amount of text in which character } t_i \text{ occurs})\}$$

$$\text{Weight } W(t_i) = IDF(t_i) \{ \sum TF(t_i, d_j) / (\text{total text amount } m) \}$$

where TF is the Term Frequency of the character t_i among a plurality d_j of documents.

Referring again to Fig. 3, when word frequency extraction in Step S3 is complete, text feature extraction is performed in Step S4. In this text feature extraction, projection onto a low-dimensional space by singular value decomposition and text feature conversion by position vectors in the low-dimensional space are performed

respectively in Steps S1 and S2, as shown by the subroutine's flowchart in Fig. 7. The processing conducted in Steps S1 and S2 corresponds to that performed by the

5 low-dimensional space projection unit 46 and the text feature conversion unit 48 included in the feature extraction unit 28 shown in Fig. 2. For low-dimensional space projection in Step 1 shown in Fig. 7, since the word-text

10 matrix 60, obtained by word frequency extraction, can be regarded as frequency vectors of individual words in each piece of text as shown in Fig. 5, each piece of text can be expressed by document vectors T_1, T_2, \dots, T_n

15 having points in a word space which is a coordinate space having words W_1, W_2, \dots, W_n as axes, as shown in Fig. 6. Therefore, the LSI (Latent Semantic Indexing) method is used to map frequency vectors of the word-text matrix

20 60 onto a space of lower dimension. The LSI method is designed to degenerate highly co-occurrent words into a single axis by singular value decomposition.

Fig. 8 is a result of singular value

25 decomposition of the word-text matrix 60 shown in Fig. 5. "p" is used as compression dimension count. Let us suppose, for example,

that $p = 20$. This singular-value-decomposed matrix illustrates a matrix 62 having word count= n along columns (n rows) and compression dimension count p for amount of text along rows (p columns),

5 a matrix 64 having compression dimension count p along both rows and columns and a matrix 66 having compression dimension count p for word count along columns and amount of text= m along rows. Of these matrices, the matrix 66

10 represents document vectors as a result of projection of vectors of text onto a low-dimensional space.

Fig. 9 represents a projection of coordinate axes C_1 and C_p (provided that $p =$

15 2), a space which is lower in dimension than document vectors W_1 to W_n in the n th order word space in Fig. 6, onto a low-dimensional space. In this low-dimensional space, since highly co-occurrent words are degenerated into a

20 single axis, the individual dimensions C_1 and C_p of the low-dimensional space correspond to a plurality of words having similar meanings, that is, combinations of a plurality of words likely to appear in a similar manner in the same

25 piece of text. For example, if high co-occurrence is observed in two words, it is possible to determine that a piece of text

containing only one of the words and another containing the other word are similar in the low-dimensional space, despite the fact that these pieces of text do not contain common words.

5 In the following Step S2 in Fig. 7, position vectors, representing the positions in the low-dimensional space as shown in Fig. 9, are extracted as text features of individual information sets.

10 Referring again to Fig. 3, when feature extraction is complete in Step S4, information sets are classified and laid out in the following Step S5. This classification and layout of information sets is conducted by
15 laying out information sets, in accordance with text-related position vectors extracted in the low-dimensional space, in a plane using the SOM (Self-Organizing Maps) as shown in Fig. 9.

20 Fig. 10 shows the procedure for classification and layout of information sets. That is, the matrix 66, obtained by text feature extraction and which will serve as a low-dimensional space, is extracted, and
25 information sets are laid out using the SOM 68 in accordance with the position vectors, and a display screen 70, on which individual images

are displayed at positions where information sets have been laid out, is created. The processing by the SOM is divided into learning and layout. The SOM's learning relies on
5 arrangement of cells in a regular manner in a plane first and then updating of each cell-assigned vector value based on the input vector value 74 assigned as a focus of learning as shown in the pre-learning map 72 in Fig. 11,
10 thereby obtaining a post-learning map 76. As a result of this learning, cells near an optimal cell 78, which has become the focus of learning 74, are laid out such that they possess similar vector values. As for the map 76, learning is
15 performed by determining a range of learning 80 as appropriate. When such a learning process is complete, the cell with a vector value closest to the target vector value is placed, based on the map 76 obtained as a result
20 of learning, at a position in the post-learning map. Consequently, highly similar pieces of text are placed at the same location, thus ensuring classification and layout in agreement with text feature.

25 Referring again to Fig. 3, when text-based classification and layout of information sets for the purpose of image

layout are complete by following Steps S1 through S5, label feature extraction and label layout are performed in Steps S6 and S7. With label feature extraction in Step S6,

5 projection onto a low-dimensional space in Step S1, label selection by significance in Step S2 and label feature conversion in Step S3 are performed as shown in the subroutine of Fig. 12. This label feature extraction for

10 classification and layout of labels can be basically conducted in the same manner as with text. First, each of the words in the word-text matrix 60, obtained by word

15 extraction as shown in Fig. 5, can be regarded as text containing only that word. Therefore, they are projected onto a low-dimensional space by singular value decomposition using the LSI method in the same manner as with text. This projection onto a low-dimensional space

20 has already been described in the text-related processing as shown in Fig. 8. Therefore, the matrix 62, having word count n and expressed by word vectors whose lengths are equal to the frequency, is extracted as the result of

25 projection onto a low-dimensional space by using text axes C_1 to C_p of the matrix 62 as coordinate axes, as shown in Fig. 13. As for

the matrix 62 having word vectors in the low-dimensional space, there are too many words to use them as labels. Therefore, label selection is performed such that only highly
5 important words are chosen. For example, a label selection 82 is performed in which 100 words are chosen in descending order of norm (significance). In this case, the lengths of word vectors in the low-dimensional space are nearly
10 equivalent to the word frequencies. Since highly important words tend to have longer word vectors in the low-dimensional space, vector lengths in the low-dimensional space are employed as significance for label selection.
15 In the following Step S3 in Fig. 12, the position vectors of selected words, obtained by projection onto the low-dimensional space, are extracted as label features.

Referring again to Fig. 3, when label
20 feature extraction in Step S6 is complete, labels are laid out in Step S7 based on label features. This label layout uses the SOM as with text. However, since labels are placed at positions which are correlated with text
25 classification and layout, the post-learning SOM 68, employed for text classification and layout, is used as is, and labels are simply

laid out with no additional learning as shown in Fig. 13, as a result of which a display screen 86 is obtained. Further, in conducting label layout, highly important labels, namely,

5 labels with longer vectors are placed toward the front in the virtual space, and labels are placed such that the more the labels are located toward the front, the larger they appear on the screen. Naturally, labels are placed at the

10 front of a predetermined two-dimensional plane in the virtual space in which text classification and layout have been performed. Consequently, those labels, placed in the

15 direction of depth as seen from the viewpoint in the virtual space, are displayed. Moreover, the more the labels are placed toward the front, the larger they appear while the more they are placed toward the back, the smaller they appear. As a result, labels existing in the direction

20 of depth and displayed in smaller size are hidden by front labels which are displayed in larger size. Therefore, it is possible to infer the contents of image information text by simply looking at the foremost label. Note

25 that as the viewpoint is moved in the direction of depth and past a certain label, that label disappears and another label behind the

vanished one appears. Note also that it is possible to determine whether to display labels in accordance with their distance from the viewpoint. By not displaying labels too close to or far from the viewpoint, it is possible to prevent an excessively close label from appearing too large or a number of small labels from being displayed together, thereby ensuring easy viewing. When one walks through a virtual space in which images and text have been laid out by moving the viewpoint, search efficiency cannot be improved unless label positions and displayed image positions are always correlated. With three-dimensional coordinate calculations, normally, the more a label is placed toward the front, the more it moves horizontally if the viewpoint is moved horizontally. Therefore, images and labels may appear displaced from each other due to the horizontal movement of the viewpoint, thus resulting in loss of their relationship. In the present invention, therefore, labels remain fixed regardless of the horizontal movement of the viewpoint, and label sizes change only with the movement of the viewpoint in the direction of depth. Thus, by not horizontally moving labels, it is possible to

move the viewpoint for walk-through while maintaining labels and images displayed in proximity to each other. When text-based classification and layout of information sets
 5 and classification and layout of labels which will be used as keywords are complete, image information and labels are displayed by information display in Step S8 of Fig. 3, as the viewpoint is moved in the virtual space with
 10 a mouse, the cursor, etc.

Figs. 14 through to 19 show changes in the screen as a user moves the viewpoint in the direction of depth. Fig. 14 shows the initial search screen in which images, classified and
 15 positioned based on text features, are displayed in a two-dimensional plane which has been placed at a predetermined position along the depth in the virtual space. At the front of these images displayed by two-dimensional
 20 layout, labels, laid out based on label feature extraction appear.

In this case, words for labels are placed such that the larger they are in size, the more they are placed toward the front. In the search
 25 screen shown in Fig. 14, suppose, for example, that the user moves the viewpoint in the direction of depth with label "鉄" as the target.

Then, label "鉄" is enlarged as shown in Fig. 15. As the viewpoint is moved further in the direction of depth toward label "鉄", label "鉄" is gradually enlarged as shown in Fig. 16.

5 Then, when the viewpoint moves past label "鉄", this label disappears and the label placed in the back appears as shown in Fig. 17. In this process, when the viewpoint is directed at label "鉄鋼" and moved further in the direction
 10 of depth, label "鉄鋼" appears enlarged as the viewpoint approaches this label as shown in Fig. 18. Moreover, when the viewpoint moves past label "鉄鋼", the image is reached which is placed at the back of the two-dimensional plane
 15 and which corresponds to label "鉄鋼" as shown in Fig. 19. Note that "鉄" and "鉄鋼" mean "iron" and "steel", respectively. When the user clicks on a specific image among those retrieved, it is possible to display text
 20 information which will serve as an information set corresponding to that image or access the web page for the retrieved information set.

Fig. 20 shows how the initial screen in Fig. 14 changes when the viewpoint, directed
 25 at label "氷" at the lower centre, is moved in the direction of depth. In this case, label "氷" disappears when the viewpoint moves past

this label. Then, the label in the back, for example, label "冰筍" comes into sight as shown in Fig. 21. Now, when the viewpoint, directed at label "冰筍", is moved further in the direction of depth and past this label, it is possible to reach a plurality of images which correspond to label "冰筍", as shown in Fig. 22. Note that "冰" and "冰筍" mean "ice" and "ice stalagmite", respectively. Thus, the multimedia information retrieval system allows desired images to be accurately and efficiently found based on labels representing the contents of text information, by conducting walk-through tours in which the viewpoint is moved by using labels placed in a virtual space as guides.

Fig. 23 depicts the use of a computer-readable record medium to store a program designed to perform multimedia information retrieval embodying this invention. That is, the program handling multimedia information retrieval

basically executes the processing steps in the flowchart of Fig. 3 by a computer. The record medium storing this program may be not only a portable record medium such as a CD-ROM or a floppy disk as shown

in Fig. 23 but could also be a record medium 106 within
a storage unit available on a network or a
record medium 108 such as a computer's hard disk
and RAM, and the program is loaded into a
5 computer 110 and run on its main memory during
program run.

Note that, in the above description,

image-text pairs are taken as
an example of information sets collected from
10 the Internet. However, such information sets
may be multimedia information further
containing moving images and music as long as
such information contains combinations of
images and text.

15 Note also that multimedia information to
be retrieved by this invention is not limited
to that available on the Internet and may be
multimedia information available on networks
or stored in servers as appropriate.

20 Note that, as for the hardware
configuration of the multimedia information
retrieval system 10 according to the present
invention, the functionality of the multimedia
information retrieval system 10 may be
25 independently provided on a server or client (e.g. PC),
or that part of the functionality may be
provided on a server with the remaining

functionality on a client.

In this case, the client need only have capacity to display image and label information placed in the virtual space, with
5 the remaining information provided on the server.

Note that, in the above description, text-feature-based classification and layout of image information are performed in a
10 predetermined two-dimensional plane in a virtual space. However, it is possible to lay out image information in the direction of depth based on text features as with labels.

Note also that this invention is not
15 limited to the specific embodiments given by way of example and includes appropriate alterations which do not impair its purpose and advantages. Further, this invention is not limited to the numbers indicated in the embodiment.

20 As described above, the present invention not only classifies and lays out information sets comprised of image-text pairs in a virtual space such that similar images are located close to each other but also allows users to
25 efficiently and properly perform visual searches for images by means of walk-through tours while grasping the contents of information by labels even if

a large number of images are laid out, since such labels, designed to make searching based on meanings and contents, are displayed, thereby ensuring efficient and highly accurate

5 multimedia information retrieval.

CLAIMS

1. A multimedia information retrieval method comprising:

5 a word frequency extraction step using as information sets paired image information and text information correlated to each other and extracting constituent word frequency information from said text information within
10 said information sets;

a text feature extraction step extracting text features, based on said word frequency information of individual information sets;

an information set classification and
15 layout step classifying and laying out said information sets in a virtual space, based on said text features;

a label feature extraction step selecting labels from constituent words of said text
20 information within each information set and extracting features of selected labels;

a label layout step placing labels at positions corresponding to information sets classified and laid out in the virtual space,
25 based on said label features; and

an information display step displaying image information and labels of said

information sets, placed in said virtual space, depending on the positions of the viewpoint.

2. The method as defined in claim 1, wherein
5 said information set classification and layout step includes classifying and laying out said information sets on a two-dimensional plane at a predetermined position in a three-dimensional virtual space, based on said
10 text features, and wherein
said label layout step includes placing said labels at the front of the two-dimensional plane in which the information sets are classified and laid out, based on said label
15 features.

3. The method as defined in claim 1 or 2, wherein
said text feature extraction step
comprises:
20 a morpheme analysis step extracting predetermined parts of speech such as nouns, noun compounds and adjectives by morpheme analyses of text information and creating a word list comprised of words used and their
25 frequencies of occurrence;

a matrix creation step creating a word-text matrix whose rows and columns

correspond respectively to text information and words and in which word frequencies of occurrence are laid out as elements; and

5 a text feature conversion step expressing text information of said word-text matrix by document vectors having coordinate points determined by frequencies of occurrence in a word space which has words as axes, projecting said document vectors onto a low-dimensional
10 space by singular value decomposition and using as text features document vectors representing positions in said low-dimensional space.

15 4. The method as defined in claim 3, wherein said matrix creation step comprises:

a weight assignment step assigning weight to elements in said word-text matrix, based on word frequencies of occurrence in each text.

20

5. The method as defined in claim 1, 2, 3 or 4, wherein said label feature extraction step comprises:

a label selection step figuring out the
25 significance of each constituent word of text information within each of said information sets and selecting words to be used as labels,

based on the significance figured out.

6. The method as defined in claim 5, wherein said label selection step comprises:
- 5 a morpheme analyses step extracting nouns and noun compounds by morpheme analyses of text information and creating a word list comprised of words used and their frequencies of occurrence;
- 10 a matrix creation step creating a word-text matrix whose rows and columns correspond respectively to text information and words and in which said word frequencies of occurrence are laid out as elements; and
- 15 a label feature conversion step expressing words of said word-text matrix by word vectors having coordinate points determined by frequencies of occurrence in a text space which has individual pieces of text
- 20 information as axes, projecting said word vectors onto a low-dimensional space by singular value decomposition and using as label features word vectors representing positions in the low-dimensional space;
- 25 wherein
- a predetermined number of words are selected as labels in descending order of

significance which is represented by lengths of said word vectors.

7. The method as defined in claim 5 or 6,
5 wherein

said label layout step includes displaying labels such that the higher the significance figured out by said label selection step, the more the labels are
10 displayed toward the front of said virtual space; and
wherein

said information display step includes changing how labels appear and the size in which
15 they appear, depending on the position of the viewpoint relative to said virtual space.

8. The method as defined in claim 7, wherein
said information display step includes
20 fixing the horizontal position of said label relative to image information regardless of a horizontal displacement of the viewpoint in said virtual space and changing how labels appear and the size in which they appear
25 depending on a change in the position of the viewpoint in the direction of depth.

9. The method as defined in any preceding claim, further comprising an information collection step collecting from the Internet information sets comprised of paired image and text information
5 correlated to each other.

10. The method as defined in claim 9, wherein said information collection step comprises:

10 a relation analysis step analyzing the relationship between image information and text information and determining the range of information to be collected as information sets, if the relationship between said image
15 information and said text information is unclear.

11. A program for retrieving multimedia information, said program allowing a computer
20 to execute:

a word frequency extraction step using as information sets paired image information and text information correlated to each other and extracting constituent word frequency
25 information from said text information within said information sets;

a text feature extraction step extracting

text features, based on said word frequency information of individual information sets;

an information set classification and layout step classifying and laying out said information sets in a virtual space, based on
5 said text features;

a label feature extraction step selecting labels from constituent words of said text information within each information set and
10 extracting features of selected labels;

a label layout step placing labels at positions corresponding to information sets classified and laid out in the virtual space, based on said label features; and

15 an information display step displaying image information and labels of said information sets, placed in said virtual space, depending on the positions of the viewpoint.

20 12. The program as defined in claim 11, wherein

said information set classification and layout step includes classifying and laying out said information sets on a two-dimensional
25 plane at a predetermined position in a three-dimensional virtual space, based on said text features, and wherein

said label layout step includes placing
said labels at the front of the two-dimensional
plane in which the information sets are
classified and laid out, based on said label
5 features.

13. The program as defined in claim 11,
wherein

said text feature extraction step
10 comprises:

a morpheme analysis step extracting
predetermined parts of speech such as nouns,
noun compounds and adjectives by morpheme
analyses of text information and creating a
15 word list comprised of words used and their
frequencies of occurrence;

a matrix creation step creating a
word-text matrix whose rows and columns
correspond respectively to text information
20 and words and in which word frequencies of
occurrence are laid out as elements; and

a text feature conversion step expressing
text information of said word-text matrix by
document vectors having coordinate points
25 determined by frequencies of occurrence in a
word space which has words as axes, projecting
said document vectors onto a low-dimensional

space by singular value decomposition and using as text features document vectors representing positions in said low-dimensional space.

5

14. The program as defined in claim 13, wherein

said matrix creation step comprises:
a weight assignment step assigning weight
10 to elements in said word-text matrix, based on word frequencies of occurrence in each text.

15. The program as defined in claim 11, wherein

15 said label feature extraction step comprises:

a label selection step figuring out the significance of each constituent word of text information within each of said information
20 sets and selecting words to be used as labels, based on the significance figured out.

16. The program as defined in claim 15, wherein

25 said label selection step comprises:
a morpheme analyses step extracting nouns and noun compounds by morpheme analyses of text

information and creating a word list comprised of words used and their frequencies of occurrence;

5 a matrix creation step creating a word-text matrix whose rows and columns correspond respectively to text information and words and in which said word frequencies of occurrence are laid out as elements; and

a label feature conversion step
10 expressing words of said word-text matrix by word vectors having coordinate points determined by frequencies of occurrence in a text space which has individual pieces of text information as axes, projecting said word
15 vectors onto a low-dimensional space by singular value decomposition and using as label features word vectors representing positions in the low-dimensional space; wherein

20 a predetermined number of words are selected as labels in descending order of significance which is represented by lengths of said word vectors.

17. The program as defined in claim 15 or 16,
25 wherein

said label layout step includes displaying labels such that the higher the

significance figured out by said label selection step, the more the labels are displayed toward the front of said virtual space; and

5 wherein

said information display step includes changing how labels appear and the size in which they appear, depending on the position of the viewpoint relative to said virtual space.

10

18. The program as defined in claim 17, wherein

said information display step includes fixing the horizontal position of said label
15 relative to image information regardless of a horizontal displacement of the viewpoint in said virtual space and changing how labels appear and the size in which they appear depending on a change in the position of the
20 viewpoint in the direction of depth.

19. The program as defined in any of claims 11 to 18, further comprising:

an information collection step
25 collecting from the Internet information sets comprised of paired image and text information correlated to each other.

20. The program as defined in claim 19,
wherein

said information collection step
5 comprises:

a relation analysis step analyzing the
relationship between image information and
text information and determining the range of
information to be collected as information
10 sets, if the relationship between said image
information and said text information is
unclear.

21. A computer readable record medium having
15 therein stored a program for retrieving
multimedia information, said program allowing
a computer to execute:

a word frequency extraction step using as
information sets paired image information and
20 text information correlated to each other and
extracting constituent word frequency
information from said text information within
said information sets;

a text feature extraction step extracting
25 text features, based on said word frequency
information of individual information sets;

an information set classification and

layout step classifying and laying out said information sets in a virtual space, based on said text features;

5 a label feature extraction step selecting labels from constituent words of said text information within each information set and extracting features of selected labels;

10 a label layout step placing labels at positions corresponding to information sets classified and laid out in the virtual space, based on said label features; and

15 an information display step displaying image information and labels of said information sets, placed in said virtual space, depending on the positions of the viewpoint.

22. A multimedia information retrieval system comprising:

20 a word frequency extraction unit using as information sets paired image information and text information correlated to each other and extracting constituent word frequency information from said text information within said information sets;

25 a text feature extraction unit extracting text features, based on said word frequency information of individual information sets;

an information set classification and layout unit classifying and laying out said information sets in a virtual space, based on said text features;

5 a label feature extraction unit selecting labels from constituent words of said text information within each information set and extracting features of selected labels;

 a label layout unit placing labels at
10 positions corresponding to information sets classified and laid out in the virtual space, based on said label features; and

 an information display unit displaying image information and labels of said
15 information sets, placed in said virtual space, depending on the positions of the viewpoint.

23. A method substantially as described herein with reference to the accompanying drawings.

20

24. A program substantially as described herein with reference to the accompanying drawings.

25. A computer readable record medium substantially
25 as described herein with reference to the accompanying drawings.

26. A multimedia information retrieval system
substantially as described herein with reference to the
30 accompanying drawings.



53



INVESTOR IN PEOPLE

Application No: GB 0227434.8
Claims searched: 1-26

Examiner: Matthew Cope
Date of search: 12 May 2003

Patents Act 1977 : Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance	
X, E	1,5,11,15, 21 & 22	EP 1262883	KODAK - Whole doc relevant
Y	1,5,11,15, 21 & 22	US 5598557	DONER - Column 3, lines 1-32 especially
Y	1,5,11,15, 21 & 22	US 6243713	NELSON - Whole doc relevant

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC^v:

G4A

Worldwide search of patent documents classified in the following areas of the IPC⁷:

G06F

The following online and other databases have been used in the preparation of this search report :

EPOQUE, WPI, PAJ

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.